



# STIC Search Report

## Biotech-Chem Library

STIC Database Tracking Number: 165971

TO: Ralph J Gitomer  
Location: 3d65 / 3c18  
Art Unit: 1655  
Friday, September 16, 2005

Case Serial Number: 10/650482

From: Noble Jarrell  
Location: Biotech-Chem Library  
Rem 1B71  
Phone: 272-2556

Noble.jarrell@uspto.gov

### Search Notes

**THIS PAGE BLANK (USPTO)**



## Protein Sequence Searches - February 2005

All of the sequence databases on ABSS have recently been updated.

- Please note that the curators of the UniProt database have purged some temporary accession numbers from the most recent version of UniProt. These sequences have been assigned new permanent accession numbers. The new UniProt record may not contain the previous temporary accession number.
- If you encounter an accession number from an older search run against UniProt (results file extension **.rup**) that can no longer be found in the database, the permanent record with the new accession number can be found by searching the old accession number in the UniProt Protein Archive database (UniPARC) at:

<http://www.pir.uniprot.org/database/archive.shtml>

If you have any questions regarding this information or your results, please contact any STIC searcher.

**When submitting sequence search results for scanning into IFW, please include a copy of this attachment to assist any future Examiners or members of the public who may encounter UniProt temporary accession numbers.**

**THIS PAGE BLANK (USPTO)**

FOR OFFICIAL USE ONLY

Scientific and Technical Information Center

SEARCH REQUEST FORM

Requester's Full Name: Ralph C. Simon Examiner #: 69630 Date: 9/16/05  
Art Unit: 1655 Phone Number: 2-0916 Serial Number: 10/650482  
Location (Bldg/Room#): 3d65 (Mailbox #): 3013 Results Format Preferred (circle): PAPER DISK  
\*\*\*\*\*

To ensure an efficient and quality search, please attach a copy of the cover sheet, claims; and abstract or fill out the following:

Title of Invention: \_\_\_\_\_

Inventors (please provide full names): \_\_\_\_\_

Earliest Priority Date: \_\_\_\_\_

Search Topic:

Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc., if known.

\*For Sequence Searches Only\* Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.

Seq 1-4 of Application  
(ALL 6ADD34L)

STAFF USE ONLY

Searcher: Noble

Searcher Phone #: \_\_\_\_\_

Searcher Location: \_\_\_\_\_

Date Searcher Picked Up: \_\_\_\_\_

Date Completed: 9/16/05

Searcher Prep & Review Time: 16

Online Time: 10

Type of Search

2 NA Sequence (#)

2 AA Sequence (#)

Structure (#)

Bibliographic

Litigation

Fulltext

Other

Vendors and cost where applicable

STN Dialog

Questel/Orbit Lexis/Nexis

Westlaw WWW/Internet

In-house sequence systems

☒ Commercial ☐ Oligomer ☐ Score/Length  
☐ Interference ☐ SPDI ☐ Encode/Transl  
Other (specify)

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 15, 2005, 08:34:33 ; Search time 12251.5 Seconds  
(without alignments)  
11635.782 Million cell updates/sec

Title: US-10-650-482-1

Perfect score: 2942

Sequence: 1 attttggcttcgctccac.....aaattatgtgctatctg 2942

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

GenEmbl.\*

1: gb\_ba.\*

2: gb\_htg.\*

3: gb\_in.\*

4: gb\_on.\*

5: gb\_ov.\*

6: gb\_pat.\*

7: gb\_ph.\*

8: gb\_pl.\*

9: gb\_pr.\*

10: gb\_ro.\*

11: gb\_sts.\*

12: gb\_sy.\*

13: gb\_un.\*

14: gb\_vi.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2934	99.7	2942	6	BD160691 Primer fo
2	2934	99.7	2942	6	AX884059 Sequence
3	2934	99.7	2942	9	AX027650 Homo sapi
4	2912.8	99.0	5264	9	BC065280 Homo sapi
5	2329	79.2	167079	9	AL606489 Human DNA
6	2329	79.2	328863	2	AL627202 Homo sapi
7	1365.6	46.4	5367	10	BC058078 Mus muscu
8	1303.4	44.3	166233	2	AC048369 Homo sapi
9	1175.4	40.0	215781	10	AC107842 Mus muscu
10	1137.4	38.7	200490	2	AC128396 Rattus no
11	1137.4	38.7	227054	2	AC098956 Rattus no
12	1137.4	38.7	248458	2	AC105827 Rattus no
13	1120.2	38.1	212261	2	AC119774 Rattus no
14	991.4	33.7	240836	2	AC094236 Rattus no
15	738	25.1	772	6	BD148615 Primer fo
16	738	25.1	772	6	AX868553 Sequence
17	618	21.0	166233	2	AC048369 Homo sapi
18	515	17.5	543	6	BD153658 Primer fo
19	515	17.5	543	6	AX873596 Sequence

c	20	413	14.0	1027	9	BC009873	BC009873 Homo sapi
c	21	287.6	9.8	348	11	G53188	G53188 SHGC-84724
c	22	265.8	9.0	561	6	AX341793	AX341793 Sequence
c	23	219	7.4	226	6	AX417281	AX417281 Sequence
c	24	219	7.4	226	6	AX977975	AX977975 Sequence
c	25	219	7.4	226	6	BD112834	BD112834 EST and e
c	26	218	7.4	218	6	AX918015	AX918015 Sequence
c	27	218	7.4	218	6	BD053548	BD053548 Sequence
c	28	207.2	7.0	3704	10	BC006897	BC006897 Mus muscu
c	29	199	6.8	706	6	AX395830	AX395830 Sequence
c	30	195	6.6	195	6	AX895150	AX895150 Sequence
c	31	195	6.6	195	6	BD030683	BD030683 Sequence
c	32	187.4	6.4	628	6	AX321354	AX321354 Sequence
c	33	61.8	2.1	2000	6	AX655393	AX655393 Sequence
c	34	61.2	2.1	125020	9	AF429315	AF429315 Homo sapi
c	35	59.4	2.0	110000	14	AV318871_2	Continuation (3 of
c	36	57.4	2.0	125020	9	AF429315	Continuation (3 of
c	37	51	1.7	7218	6	I66494	I66494 Sequence 14
c	38	50.4	1.7	110000	3	AC116305_3	Continuation (4 of
c	39	50	1.7	60360	5	CR391988	CR391988 Zebrafish
c	40	48.6	1.7	110000	3	AC116957_2	Continuation (3 of
c	41	48.6	1.7	162778	9	AC009559	AC009559 Homo sapi
c	42	48	1.6	140982	5	BX842684	BX842684 Zebrafish
c	43	47.6	1.6	245459	2	AC016186	AC016186 Homo sapi
c	44	47.2	1.6	153445	2	AC143017	AC143017 Macaca mu
c	45	46.4	1.6	8346	6	AX323714	AX323714 Sequence
c	46	46.2	1.6	4001	6	CQ787366	CQ787366 Sequence
c	47	46.2	1.6	167320	5	AL928691	AL928691 Zebrafish
c	48	46.2	1.6	17467	9	AC097510	AC097510 Homo sapi
c	49	46	1.6	172697	10	AC123849	AC123849 Mus muscu
c	50	46	1.6	199698	2	BX004991	BX004991 Danio rer
c	51	45.8	1.6	149464	2	AC067800	AC067800 Homo sapi
c	52	45.8	1.6	158811	9	AC004862	AC004862 Homo sapi
c	53	45.4	1.5	1141	6	AX083744	AX083744 Sequence
c	54	45.4	1.5	228767	5	BX005261	BX005261 Zebrafish
c	55	45	1.5	250029	3	AE014820	AE014820 Plasmodi
c	56	44.8	1.5	4001	6	CQ787370	CQ787370 Sequence
c	57	44.8	1.5	200953	2	AC130578	AC130578 Rattus no
c	58	44.8	1.5	253275	2	AC130147	AC130147 Rattus no
c	59	44.8	1.5	262008	2	AC129449	AC129449 Rattus no
c	60	44.6	1.5	143096	2	CR387991	CR387991 Danio rer
c	61	44.2	1.5	39575	3	AC116925	AC116925 Dictyoste
c	62	44.2	1.5	108175	9	AC068206	AC068206 Homo sapi
c	63	44.2	1.5	153264	2	AC135216	AC135216 Bos tauru
c	64	44.2	1.5	162974	2	AC046166	AC046166 Homo sapi
c	65	44.2	1.5	167478	2	AC067723	AC067723 Homo sapi
c	66	44.2	1.5	169571	9	AC020687	AC020687 Homo sapi
c	67	44.2	1.5	190102	9	AF003085	AF003085 Homo sapi
c	68	44.2	1.5	191654	2	AC023176	AC023176 Homo sapi
c	69	44.2	1.5	209152	2	AC015558	AC015558 Homo sapi
c	70	44	1.5	166850	2	AC012222	AC012222 Homo sapi
c	71	44	1.5	175346	2	AC130426	AC130426 Homo sapi
c	72	44	1.5	300684	1	AE017227	AE017227 Mycobacte
c	73	43.8	1.5	1051	9	CHRGIT	M30947 Chimpancee
c	74	43.8	1.5	163495	5	BX321919	BX321919 Zebrafish
c	75	43.8	1.5	178050	9	AC142324	AC142324 Pan trogl
c	76	43.6	1.5	258144	2	BX927403	BX927403 Danio rer
c	77	43.6	1.5	196685	2	AC092517	AC092517 Papio anu
c	78	43.4	1.5	5986	6	AX323783	AX323783 Sequence
c	79	43.4	1.5	200418	9	AC074138	AC074138 Homo sapi
c	80	43.4	1.5	200688	2	AC073266	AC073266 Homo sapi
c	81	43.2	1.5	6794	6	AX251873	AX251873 Sequence
c	82	43.2	1.5	6794	6	AX344261	AX344261 Sequence
c	83	43.2	1.5	6794	6	AX348652	AX348652 Sequence
c	84	43.2	1.5	11026	1	AE014120	AE014120 Buchnera
c	85	43.2	1.5	155902	9	AF002852	AF002852 Homo sapi
c	86	43.2	1.5	176734	2	CR377214	CR377214 Danio rer
c	87	43.2	1.5	184332	2	AC011777	AC011777 Homo sapi
c	88	43.2	1.5	215351	10	AL844212	AL844212 Mouse DNA
c	89	43	1.5	37447	3	CEF11A5	292830 Caenorhabdi
c	90	43	1.5	122163	8	AC125793	AC125793 Oryza sat
c	91	42.8	1.5	137156	2	AC138846	AC138846 Homo sapi
c	92	42.8	1.5	148255	10	AC121524	AC121524 Mus muscu

c 93	42.8	1.5 154803	9	AC131392	AC131392 Homo sapi
c 94	42.8	1.5 160901	2	AC138812	AC138812 Homo sapi
c 95	42.8	1.5 161000	2	AC145100	AC145100 Homo sapi
c 96	42.8	1.5 162270	2	AC140170	AC140170 Homo sapi
c 97	42.8	1.5 164575	2	AC024736	AC024736 Homo sapi
c 98	42.8	1.5 165211	2	AC145128	AC145128 Homo sapi
c 99	42.8	1.5 165867	2	AC145105	AC145105 Homo sapi
c 100	42.8	1.5 167663	2	AC145135	AC145135 Homo sapi
ALIGNMENTS					
RESULT 1	BD160691 2942 bp DNA linear PAT 17-JAN-2003				
LOCUS	Primer for synthesizing full-length cDNA and use thereof.				
DEFINITION	BD160691				
ACCESSION	BD160691				
VERSION	BD160691.1 GI:27866449				
KEYWORDS	JP 2002191363-A/15534.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
TITLE	Ota,T., Isogai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J.,				
JOURNAL	Ishii,S., Sugiyama,T., Wakamatsu,A., Nagai,K. and Otsuki,T.				
COMMENT	Primer for synthesizing full-length cDNA and use thereof				
FT CDS	Patent: JP 2002191363-A 15534 09-JUL-2002;				
source	HELIX RESEARCH INSTITUTE				
OS	Homo sapiens (human)				
PN	JP 2002191363-A/15534				
PD	09-JUL-2002				
PI	28-JUL-2000 JP 2000280990				
PI	TOSHIO OPA, TAKAO ISOGAI, TETSUO NISHIKAWA, KOJI HAYASHI, KAORU				
PI	SAITO,				
PI	JUNICHI YAMAMOTO, SHIZUKO ISHII, TOMOYASU SUGIYAMA, AI WAKAMATSU,				
PI	KEIICHI NAGAI, TETSUJI OTSUKI				
PC	C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/10,				
PC	C12P21/02, C12Q1/68//C12P21/08, G06F17/30, C12N15/00, C12N5/00 CC				
Key	Primer for synthesizing full-length cDNA and use thereof				
Location/Qualifiers	(407) . . (2545).				
FT CDS	Location/Qualifiers				
source	1. .2942				
ORIGIN	/organism="Homo sapiens"				
Query Match	99.7%; Score 2934; DB 6; Length 2942;				
Best Local Similarity	99.8%; Pred. No. 0;				
Matches 2937; Conservative	0; Mismatches 5; Indels 0; Gaps 0;				
QY	1	ATTTTGGGCTCGCTTCCACGCGCACGCGGCGCTACCCAGTCCCTTCGCGTATCGCGTTG	60		
Db	1	ATTTTGGGCTTCGCTTCCACCGCACGCGGCGCTACCCAGTCCCTTCGCGTATCGCGTTG	60		
QY	61	CTCAGGGGCTTTCAACCTCTGTGTCAGTCGGAACCACTCGCGGAGGCGGTGGGGGACT	120		
Db	61	CTCAGGGGCTTTCAACCTCTGTGTCAGTCGGAACCACTCGCGGAGGCGGTGGGGGACT	120		
QY	121	CCTATCCATGGTGTGAACGTCGAGCGGACCTAGGGAACTCTTCCCGCGCAGGATGGA	180		
Db	121	CCTATCCATGGTGTGAACGTCGAGCGGACCTAGGGAACTCTTCCCGCGCAGGATGGA	180		
QY	181	AGTGCATCAGTCGCGCGCTATTGGCGGGTGTTCCTCCCTGTGCTTCGCGCCGCTG	240		
Db	181	AGTGCATCAGTCGCGCGCTATTGGCGGGTGTTCCTCCCTGTGCTTCGCGCCGCTG	240		
QY	241	CGCATTCGCTGCCCTCTGTGGCTTTTCTGTGGCTCGAAGATCGGCGCTGGAGCGAC	300		

Db	241	CCGCAATTCGCTGCCCTCTGTGGCTTTTCTGTGGCTCGAAGATCGGCGCTGGAGCGGAC	300		
QY	301	GCACCGCTGGGGAAGCGCGAGACTCTGTAGGCTTCTCGAATCCCGTCCAGCTCCAGC	360		
Db	301	GCACCGCTGGGGAAGCGCGAGACTCTGTAGGCTTCTCGAATCCCGTCCAGCTCCAGC	360		
QY	361	CGCTGAGCGCGCGGCTTACCTGAGAGACTGTCAAGAAAAAGAGATGGAGCGGGGAC	420		
Db	361	CGCTGAGCGCGCGGCTTACCTGAGAGACTGTCAAGAAAAAGAGATGGAGCGGGGAC	420		
QY	421	AGGCGGATCGCGGAAAAGGCTTGGCCCTCGGCGGGGCTTCCGGTTCGCGCAACCTTTT	480		
Db	421	AGGCGGATCGCGGAAAAGGCTTGGCCCTCGGCGGGGCTTCCGGTTCGCGCAACCTTTT	480		
QY	481	CCCTCGGGATCGCAAGCAGGCTCTTCTAAGTTCCCGACGCTCTTCCGCGGAAACTC	540		
Db	481	CCCTCGGGATCGCAAGCAGGCTCTTCTAAGTTCCCGACGCTCTTCCGCGGAAACTC	540		
QY	541	CGGGAACCCACACTGCTTTCCTGCGCCAGCCGAGACTCGGCTCAGTTACTGGACGAA	600		
Db	541	CGGGAACCCACACTGCTTTCCTGCGCCAGCCGAGACTCGGCTCAGTTACTGGACGAA	600		
QY	601	ACTGCTCTCCAGCTCCTTGGCGGCTCCCGGATTCCTCAGAGGTGCTTAATTTGGAG	660		
Db	601	ACTGCTCTCCAGCTCCTTGGCGGCTCCCGGATTCCTCAGAGGTGCTTAATTTGGAG	660		
QY	661	CCAACTTTTGGTGGAAATGTTCCGACAGATGCTAGATTTTCTGAGGCTCTACAGCGC	720		
Db	661	CCAACTTTTGGTGGAAATGTTCCGACAGATGCTAGATTTTCTGAGGCTCTACAGCGC	720		
QY	721	CCTGAGAGCCCTGAAGGGAACCGAGAAACCGCGCCGCCACAGCGCAGAAATCTTTGAG	780		
Db	721	CCTGAGAGCCCTGAAGGGAACCGAGAAACCGCGCCGCCACAGCGCAGAAATCTTTGAG	780		
QY	781	TTGCTCGAGCTCGACTCTCCTCAGAGCCCTCGGTCCAGTCCCTTGAATGGCTAGAGGA	840		
Db	781	TTGCTCGAGCTCGACTCTCCTCAGAGCCCTCGGTCCAGTCCCTTGAATGGCTAGAGGA	840		
QY	841	GGGATCCACTGGCAATCTCGCCCGCAGACCTAAATTTGGAGCTTAAGCCAGGGAAG	900		
Db	841	GGGATCCACTGGCAATCTCGCCCGCAGACCTAAATTTGGAGCTTAAGCCAGGGAAG	900		
QY	901	TGCTTTGGAGCCCTGACGACAGGCTTTTCTTTAGAGCAGCAGCTGTGGGAGTGGAGCT	960		
Db	901	TGCTTTGGAGCCCTGACGACAGGCTTTTCTTTAGAGCAGCAGCTGTGGGAGTGGAGCT	960		
QY	961	GTTCGCCAGTAGCTTCAATCCCGTCTGTACTCTTAACCGGGAACTTGGCTTCGCGCTC	1020		
Db	961	GTTCGCCAGTAGCTTCAATCCCGTCTGTACTCTTAACCGGGAACTTGGCTTCGCGCTC	1020		
QY	1021	TGGGCTCTAAACATTCACGCGATAGACAAATTTTCAGTGTGTATCTTATTTGCTGAACCC	1080		
Db	1021	TGGGCTCTAAACATTCACGCGATAGACAAATTTTCAGTGTGTATCTTATTTGCTGAACCC	1080		
QY	1081	TTCTACTCTGAGCTGCTTTTCTTAGGCTAGAAGTCAAGTATCAGAACAGTGAATAG	1140		
Db	1081	TTCTACTCTGAGCTGCTTTTCTTAGGCTAGAAGTCAAGTATCAGAACAGTGAATAG	1140		
QY	1141	CGAGGTAGTGGCTTCAGACACTAACCCAGAGAGCAGCTGCTGAGAGAGACCAATG	1200		
Db	1141	CGAGGTAGTGGCTTCAGACACTAACCCAGAGAGCAGCTGCTGAGAGAGACCAATG	1200		
QY	1201	TCATCCCGCGCTGAGTCGAGAACTCATTCGCGCTTCGCGAGGAGTGCACCTCT	1260		
Db	1201	TCATCCCGCGCTGAGTCGAGAACTCATTCGCGCTTCGCGAGGAGTGCACCTCT	1260		
QY	1261	TTCTACGGAAGGCTACAGAAATTCACCATCTTCGATGAAAGCGGTGGAATTCCTTCA	1320		
Db	1261	TTCTACGGAAGGCTACAGAAATTCACCATCTTCGATGAAAGCGGTGGAATTCCTTCA	1320		
QY	1321	ACAGGCTTAAACAGGGGCAAGATTTACCCACCTTCAGCAGGATTAATGGCTACACAGCT	1380		

1321	ACAGGCTAGCAAGGGGCAAGATTATTA	CCCA	CCCCCTGAC	CCAGGATAATGGCTAC	CA	CAGCCT	1380
1381	GGAGGGAGAA	CACAGACCCCTTCTCGGATG	GCATCA	AAAA	CAC	TGCAGAGATAAC	1440
1381	GGAGGGAGAA	CACAGACCCCTTCTCGGATG	GCATCA	AAAA	CAC	TGCAGAGATAAC	1440
1441	GTTTGTTCTCTGCTGGAGACATTC	TCTG	GA	AAAA	CAC	CCAGGAATCC	1500
1441	GTTTGTTCTCTGCTGGAGACATTC	TCTG	GA	AAAA	CAC	CCAGGAATCC	1500
1501	AGAAATTATTAATACAGAGGTTCC	AC	TTC	TG	GA	AGAGAGGCGCTTCTCAGGGCTG	1560
1501	AGAAATTATTAATACAGAGGTTCC	AC	TTC	TG	GA	AGAGAGAGGCGCTTCTCAGGGCTG	1560
1561	TCCATCTAGTGAGATACCTAT	TG	AAAA	AGGAGCC	TGG	AGAGGGCCGAATAA	1620
1561	TCCATCTAGTGAGATACCTAT	TG	AAAA	AGGAGCC	TGG	AGAGGGCCGAATAA	1620
1621	TTACTCTATACCTAGAGGTCACCT	TCC	AT	TTCTG	CC	AGACCAAGCTTGTAGTAACAACT	1680
1621	TTACTCTATACCTAGAGGTCACCT	TCC	AT	TTCTG	CC	AGACCAAGCTTGTAGTAACAACT	1680
1681	GATAGATTAATATTTGGGAGGTC	ATCC	AG	TG	AA	CAAGTTC	1740
1681	GATAGATTAATATTTGGGAGGTC	ATCC	AG	TG	AA	CAAGTTC	1740
1741	TGAGGATTGGGATGAGGAAGCT	GAG	AT	GT	TTT	TGATAGTAC	1800
1741	TGAGGATTGGGATGAGGAAGCT	GAG	AT	GT	TTT	TGATAGTAC	1800
1801	CTCAGACCTTTGAA	CA	AG	CCCTG	AA	GGGCTTACCTTTG	1860
1801	CTCAGACCTTTGAA	CA	AG	CCCTG	AA	GGGCTTACCTTTG	1860
1861	TCCTTATATCCCCAGABACTT	TA	C	AG	CA	AAATTC	1920
1861	TCCTTATATCCCCAGABACTT	TA	C	AG	CA	AAATTC	1920
1921	AGAGCCTTCTGATTC	C	AG	AA	GAAGATTTGT	CTGGCAAGCTCTG	1980
1921	AGAGCCTTCTGATTC	C	AG	AA	GAAGATTTGT	CTGGCAAGCTCTG	1980
1981	GTCTGGAAGCCTTCTCTG	AG	AC	TG	ATGTTCT	GGGAGGAAGATG	2040
1981	GTCTGGAAGCCTTCTCTG	AG	AC	TG	ATGTTCT	GGGAGGAAGATG	2040
2041	TAGTGACAGATGA	AG	C	AG	AGATCT	CAAACTCTG	2100
2041	TAGTGACAGATGA	AG	C	AG	AGATCT	CAAACTCTG	2100
2101	CTACAACCCCTTTAAATTTT	AA	GGCTC	CTTTT	CA	AA	2160
2101	CTACAACCCCTTTAAATTTT	AA	GGCTC	CTTTT	CA	AA	2160
2161	TCGTGACTCA	AA	AG	ACCCCAT	TG	TG	2220
2161	TCGTGACTCA	AA	AG	ACCCCAT	TG	TG	2220
2221	TTCTTTGTAAGGTG	C	AG	CTGTT	GG	AG	2280
2221	TTCTTTGTAAGGTG	C	AG	CTGTT	GG	AG	2280
2281	TGACGTTCTTTCTG	AG	GA	AGAC	C	A	2340
2281	TGACGTTCTTTCTG	AG	GA	AGAC	C	A	2340
2341	AGAAGTTACT	C	AG	TA	TATAT	AG	2400
2341	AGAAGTTACT	C	AG	TA	TATAT	AG	2400
2401	TGCAAGGGATG	G	A	TG	CT	CA	2460
2401	TGCAAGGGATG	G	A	TG	CT	CA	2460

QY	2461	TTGCTTGACATTTTGAAACACAGAGAAAGATGTTTAAATAGACTCCAGGGAACTGCTTCAA	2520
Db	2461	TTGCTTGACATTTTGAAACACAGAGAAAGATGTTTAAATAGACTCCAGGGAACTGCTTCAA	2520
QY	2521	AGGACTTAAATGTTCTCAAGCAATGTTGAGTTGGCAGGCTGTAGTCTCTAGCTAGCATACAC	2580
Db	2521	AGGACTTAAATGTTCTCAAGCAATGTTGAGTTGGCAGGCTGTAGTCTCTAGCTAGCATACAC	2580
QY	2581	TACCTCTTAACCTGAGAGGTTGCTTTTAAAAACAAATCTTGCGCAGCTGCTTTTGA	2640
Db	2581	TACCTCTTAACCTGAGAGGTTGCTTTTAAAAACAAATCTTGCGCAGCTGCTTTTGA	2640
QY	2641	TTTTTTTTTAGAGGAAATGTAACCTGGATCTAGTTTTAAATTTTTTTTTTTTGGCAACATATCCC	2700
Db	2641	TTTTTTTTTAGAGGAAATGTAACCTGGATCTAGTTTTAAATTTTTTTTTTTTGGCAACATATCCC	2700
QY	2701	ACTCAGAAAACATTCAGGCTTTGAAGCCAGCCCTGATATGAAGGATGAACCTAGTGTGATTT	2760
Db	2701	ACTCAGAAAACATTCAGGCTTTGAAGCCAGCCCTGATATGAAGGATGAACCTAGTGTGATTT	2760
QY	2761	CTAATCCTCCCTTTTTTTGATTTAGTTGGATGTGCTTTTTTAAATGTCTTTCGCTGCATGAG	2820
Db	2761	CTAATCCTCCCTTTTTTTGATTTAGTTGGATGTGCTTTTTTAAATGTCTTTCGCTGCATGAG	2820
QY	2821	GTGGAAAGGGGACCTTTTTTCAGTGTGTCATTTTGCACATTCAGAACTATTTTCTTGAAA	2880
Db	2821	GTGGAAAGGGGACCTTTTTTTGAGTGTGTCATTTTGCACATTCAGAACTATTTTCTTGAAA	2880
QY	2881	ACAATATTTATAGGCTTTAAAGCCCATTTTCATTTCTAAATCTAAATATGTGTGCCTATC	2940
Db	2881	ACAATATTTATAGGCTTTAAAGCCCATTTTCATTTCTAAATCTAAATATGTGTGCCTATC	2940
QY	2941	TG 2942	
Db	2941	TG 2942	
RESULT 2			
AX884059		2942 bp	DNA
LOCUS	AX884059	18964	from Patent EP1074617.
DEFINITION	Sequence		
ACCESSION	AX884059		
VERSION	AX884059.1	GI:40038960	
KEYWORDS			
SOURCE	Homio sapiens (human)		
ORGANISM	Homio sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homioidae; Homo.		
AUTHORS	Ota, T., Isoqai, T., Nishikawa, T., Hayashi, K., Saito, K., Yamamoto, J., Ishii, S., Sugiyama, T., Wakamatsu, A., Nagai, K. and Otsuki, T.		
TITLE	Primers for synthesising full-length cDNA and their use		
JOURNAL	Patent: EP 1074617-A 18964 07-FEB-2001;		
FEATURES	Research Association for Biotechnology (JP)		
source	Location/Qualifiers		
	1. .2942		
	/organism="Homio sapiens"		
	/mol_type="unassigned DNA"		
	/db_xref="taxon:9606"		
	407. .2548		
	/note="unnamed protein product"		
	/codon_start=1		
	/protein_id="CA892217.1"		
	/db_xref="GI:40038961"		
	/translation="MEPTLSSQREKPAAPPAOKSLASLQDSSDVTSPLDWLEEGIHWOYSPDCLKLE		
	YSALRALKGREKPAAPPAOKSLASLQDSSDVTSPLDWLEEGIHWOYSPDCLKLE		
	KAGSALDPAQAQAFLEQQWQVLEPSSLSQSLYSNRELGGSSGCLNQRIDDFSV		
	VSYLNPSSLDLDFRLEVSQNSDGNSEVGFQTLTPESCLRDHCHPOPLAELLIP		
	ASWQCGPSTEGTEPLTHLMKLELEFLOQASQDLPDQDNGHSLBEEHSLRLM		
	DPKHGRNPQFPAAGDIQDNPQTEESKIELTTVEPLALEEESPECEGSEIPM		
	EKBPGEGRNISVDYSYLEGDLPIASAPACSNKLIDYILGASDLETSPEGEQMDWE		
CDS			

[illegible]



```
Oy 2041 TAGTCAGATGAAGCAGAGAGTCTCAAACTGTGGAACCTCATTTCTGTAAATCTGATGACCC 2100
Db 2041 TAGTCAGATGAAGCAGAGAGTCTCAAACTGTGGAACCTCATTTCTGTAAATCTGATGACCC 2100

Oy 2101 CTACAACCCCTTAAATTTTAAGGCTCTCTTTCAAAACATCAGGGGAAAATGAGAAAGGCTG 2160
Db 2101 CTACAACCCCTTAAATTTTAAGGCTCTCTTTCAAAACATCAGGGGAAAATGAGAAAGGCTG 2160

Oy 2161 TCGTGACTCAAAGACCCCATCTGAGTGCATTTCTGTGGCCATTTCTCAGTGTGCACACCTTACT 2220
Db 2161 TCGTGACTCAAAGACCCCATCTGAGTGCATTTCTGTGGCCATTTCTCAGTGTGCACACCTTACT 2220

Oy 2221 TTCTTTGAAGGTGAGCTGTTGGGGAGCCAGAAAGTGAATGTCCAGACTCGGTACAGCG 2280
Db 2221 TTCTTTGAAGGTGAGCTGTTGGGGAGCCAGAAAGTGAATGTCCAGACTCGGTACAGCG 2280

Oy 2281 TGACGTTCTTCTCGGAGACACACACATGTCAAAAGAAAAGGTAACCTTCTCTTGA 2340
Db 2281 TGACGTTCTTCTCGGAGACACACACATGTCAAAAGAAAAGGTAACCTTCTCTTGA 2340

Oy 2341 AGAAGTTACTGAGTATTATATAAGTGTGTGATGAGGATCGCAAGGACCATGGGAAGATT 2400
Db 2341 AGAAGTTACTGAGTATTATATAAGTGTGTGATGAGGATCGCAAGGACCATGGGAAGATT 2400

Oy 2401 TGCAAGGGATGGATGCGAGGTTCCAGAAACGAAATTCAGAAACAGAGATGCTATTGGATA 2460
Db 2401 TGCAAGGGATGGATGCGAGGTTCCAGAAACGAAATTCAGAAACAGAGATGCTATTGGATA 2460

Oy 2461 TTGCTTGACATTTGAAACACAGAGAAAGATGTTTAAATAGACTCCAGGGAACATGCTTCAA 2520
Db 2461 TTGCTTGACATTTGAAACACAGAGAAAGATGTTTAAATAGACTCCAGGGAACATGCTTCAA 2520

Oy 2521 AGGACTTAATGTTCTCAAGCAATGTTGAGTTGGCAGCTGTAGTCTAGCTAGCATACAC 2580
Db 2521 AGGACTTAATGTTCTCAAGCAATGTTGAGTTGGCAGCTGTAGTCTAGCTAGCATACAC 2580

Oy 2581 TACCTCTTACCTGAGAGGTGTCTTTTAAACAAATCTTGGCAGCTGTCTTTGACATTT 2640
Db 2581 TACCTCTTACCTGAGAGGTGTCTTTTAAACAAATCTTGGCAGCTGTCTTTGACATTT 2640

Oy 2641 TTTTCTTTAGAGGAAATGTAATCTGGATCTAGTTAAATTTTTTTTTTTTGGCAATATCCC 2700
Db 2641 TTTTCTTTAGAGGAAATGTAATCTGGATCTAGTTAAATTTTTTTTTTTTGGCAATATCCC 2700

Oy 2701 ACTCAGAAACATTCAGGTTTGAAGCCGCTGATGATGAAGGATGAAGTGTGATTT 2760
Db 2701 ACTCAGAAACATTCAGGTTTGAAGCCGCTGATGATGAAGGATGAAGTGTGATTT 2760

Oy 2761 CTAACTCTCCCTTTTGTAGTTAGTGTGCTTTTAAATGTCTTTGCTGCTGATGAG 2820
Db 2761 CTAACTCTCCCTTTTGTAGTTAGTGTGCTTTTAAATGTCTTTGCTGCTGATGAG 2820

Oy 2821 GTGAAAGGGGACCTTTTGTAGTTGTCAATTTTGGACATTTTCAAACTTATTTCTTGAAA 2880
Db 2821 GTGAAAGGGGACCTTTTGTAGTTGTCAATTTTGGACATTTTCAAACTTATTTCTTGAAA 2880

Oy 2881 ACAATATTATAGGCTTAAAGCCCATTTTCATTTCTAACTAAATATGTCGCTTATC 2940
Db 2881 ACAATATTATAGGCTTAAAGCCCATTTTCATTTCTAACTAAATATGTCGCTTATC 2940

Oy 2941 TG 2942
Db 2941 TG 2942
```

RESULT 3  
LOCUS AK027650  
DEFINITION Homo sapiens cDNA FLJ14744 fis, clone NT2RP3002770.  
ACCESSION AK027650  
VERSION AK027650.1  
KEYWORDS AK027650.1 Gr:14042484  
oligo capping; fis (full insert sequence).

SOURCE  
ORGANISM

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

## REFERENCE

## AUTHORS

1  
Ota, T., Suzuki, Y., Nishikawa, T., Otsuki, T., Sugiyama, T., Irie, R.,  
Wakamatsu, A., Hayashi, K., Sato, H., Nagai, K., Kimura, K., Makita, H.,  
Sekine, M., Ohtsuka, M., Nishi, T., Shibahara, T., Tanaka, T.,  
Ishii, S., Yamamoto, J., Saito, K., Kawai, Y., Isono, Y., Nakamura, Y.,  
Nagahari, K., Murakami, K., Yasuda, T., Iwayanagi, T., Wagatsuma, M.,  
Shiratori, A., Sudo, H., Hosoiri, T., Kaku, Y., Kodaira, H., Kondo, H.,  
Sugawara, M., Takahashi, M., Kanda, K., Kanda, K., Yokoi, T., Furuya, T.,  
Kikkawa, E., Omura, Y., Abe, K., Kamihara, K., Katsuta, N., Sato, K.,  
Tanikawa, M., Yamazaki, M., Ninomiya, K., Ishibashi, T., Yamashita, H.,  
Murakawa, K., Fujimori, K., Tanai, H., Kimata, M., Watanabe, M.,  
Hiraoka, S., Chiba, Y., Ishida, S., Ono, Y., Takiguchi, S., Watanabe, S.,  
Yosida, M., Hotuta, T., Kusanagi, J., Kanehori, K., Takahashi-Fujii, A.,  
Hara, H., Tanase, T., Nomura, Y., Togiya, S., Komai, F., Hara, R., Sato, K.,  
Takeuchi, K., Arita, M., Imose, N., Musashino, K., Yuuki, H., Oshima, A.,  
Sasaki, N., Aotsuka, S., Yoshikawa, Y., Matsunawa, H., Ichihara, T.,  
Shiohata, N., Sano, S., Moriya, S., Momiyama, H., Satoh, N., Takami, S.,  
Teraashima, Y., Suzuki, O., Nakagawa, S., Senoh, A., Mizoguchi, H.,  
Goto, Y., Shimizu, F., Wakebe, H., Hishigaki, H., Watanabe, T.,  
Sugiyama, A., Takemoto, M., Kawakami, B., Yamazaki, M., Watanabe, K.,  
Kumagai, A., Itakura, S., Fukuzumi, Y., Fujimori, Y., Komiyama, M.,  
Tashiro, H., Tanigami, A., Fujiwara, T., Ono, T., Yamada, K., Fujii, Y.,  
Ozaki, K., Hirao, M., Ohmori, Y., Kawabata, A., Hikiji, T., Kobatake, N.,  
Inagaki, H., Ikema, Y., Okamoto, S., Okitani, R., Kawakami, T.,  
Noguchi, S., Itoh, T., Shigeta, K., Senba, T., Matsumura, K.,  
Nakajima, Y., Mizuno, T., Morinaga, M., Sasaki, M., Togashi, T.,  
Oyama, M., Hata, H., Watanabe, M., Komatsu, T., Mizushima-Sugano, J.,  
Sato, T., Shirai, Y., Takahashi, Y., Nakagawa, K., Okumura, K.,  
Nagase, T., Nomura, N., Kikuchi, H., Masuho, Y., Yamashita, R.,  
Nakai, K., Yada, T., Nakamura, Y., Ohara, O., Isogai, T. and Sugano, S.  
Complete sequencing and characterization of 21,243 full-length  
human cDNAs  
Nat. Genet. 36 (1), 40-45 (2004)  
14702039

## TITLE

## JOURNAL

## PUBMED

## REFERENCE

## AUTHORS

2  
Isogai, T., Ota, T., Hayaashi, K., Sugiyama, T., Otsuki, T., Suzuki, Y.,  
Nishikawa, T., Nagai, K., Sugano, S., Takahashi-Fujii, A., Hara, H.,  
Tanase, T., Nomura, Y., Togiya, S., Komai, F., Hara, R., Takeuchi, K.,  
Arita, M., Nabekura, T., Ishii, S., Kawai, Y., Saito, K., Yamamoto, J.,  
Wakamatsu, A., Nakamura, Y., Nagahari, K., Masuho, Y. and Oshima, A.  
NEDO human cDNA sequencing project  
Unpublished  
3 (bases 1 to 2942)  
Isogai, T. and Otsuki, T.  
Direct Submission

## JOURNAL

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

COMMENT  
Economy, Trade and Industry of Japan; cDNA full insert sequencing:  
Research Association for Biotechnology; cDNA library construction,  
5'- 3'-end one pass sequencing and clone selection: Helix  
Research Institute (supported by Japan Key Technology Center etc.)  
and Department of Virology, Institute of Medical Science,  
University of Tokyo.

## FEATURES

## source

1. 2942  
/location="Qualifiers  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="NT2RP3002770"  
/cell\_line="NT2"  
/cell\_type="teratocarcinoma"  
/clone\_lib="NT2RP3"  
/notes="cloning vector: pME18SFL3-mRNA from NT2 neuronal  
precursor cells after 2-weeks retinoic acid (RA)  
induction."  
407. 2548  
/note="unnamed protein product"

## CDS

/codon_start=1 /protein_id="BAB55266.1" /db_xref="GI:14042485" /translation="MEPTGSGSRKRLGPRAGRFWPPFPFRSQSGSKFTPTPLGPN SGNPTLLSAGPETHVTSWTKLLSOLLAPLGLLQKVLINLSQLFQGMFPTRWLDFAVY YSALRALGRKPAAPTAKQLSLSDSPSVTSPLDMLBEGHWHQYSPPDLKJEL KAKGSALPAQAFLLEQLMGVELLPSLQSLRYLNRELGSFSPGLNIQRIIDFVS VSYLLNPVLDLCPFLREYSYONSGNSVWGFTLTPSSCLRDHCHPOPLSABLIP ASWQCPPLSTEGELPEIHLRMKRLFLQQAASKQDLPTPDQNGYHSLREHSILRM DPKCRDNPTEVPAAGDI PGNTQBSTBEKIELLTTEVPLALBESPSGSPSEI PM EKPEGRHSVVDYSLEGDLPISARPACSNKLIDYILGASSDLETSDPEGEDWDE RAEDQFDSLSDSLEQPEGLHLNFSVDPYNPONTATIQTAAIRIVPEPS DSEKLDGSKDLENSQSGSLPETPEHSGSEEDWESSADEAESLKLMSFNCSDDPY NPLNPKA POTS GENEKCRDSKTPSESI VAI SECHTLLSKVOLLGSQESCEPDSVQ RDVLSGSHTHVKRKKVTFLEVTYVYISGEDRKG PWEERFARDGCRFPQKRIQETEDA IGYCLTPEHREMFNRLOGTCFKGLNLVKQC"									
ORIGIN									
Query Match 99.7%; Score 2934; DB 9; Length 2942; Best Local Similarity 99.8%; Pred. No. 0; Matches 2937; Conservative 0; Mismatches 5; Indels 0; Gaps 0;									
Qy	1	ATT	TGGGCTTCG	TTCCACCCGAC	CCAGCGGCTTACCCAGTCTTC	CGGTATCGG	TTG	60	
Db	1	ATT	TTGGGCTTCG	TTCCACCCGAC	CCAGCGGCTTACCCAGTCTTC	CGGTATCGG	TTG	60	
Qy	61	CTC	AGGGGCTTTT	CAACCTCTGT	CAGTCGGAACCACTCGCCGAGGCGCTGGGGGACT	120			
Db	61	CTC	AGGGGCTTTT	CAACCTCTGT	CAGTCGGAACCACTCGCCGAGGCGCTGGGGGACT	120			
Qy	121	CCT	ATCCATGTGT	TTGAAGCGT	CGAGCGGACTTAGGGAACCTCTTCCCGCCGACGAGATGGA	180			
Db	121	CCT	ATCCATGTGT	TTGAAGCGT	CGAGCGGACTTAGGGAACCTCTTCCCGCCGACGAGATGGA	180			
Qy	181	AGT	CGGATCAGT	CGCGGCTAT	TGCGGGGCTGTGTTCTTCCCTGTGTTCTGCGCCCGCTG	240			
Db	181	AGT	CGCATCAGT	CGCGGCTAT	TGCGGGGCTGTGTTCTTCCCTGTGTTCTGCGCCCGCTG	240			
Qy	241	CGC	ATTTCGTC	CGCCCTCTGT	GTGGCTTTTCTGTGGCTCGAAGATCGGCTCGAGACGCGAC	300			
Db	241	CGC	ATTTCGTC	CGCCCTCTGT	GTGGCTTTTCTGTGGCTCGAAGATCGGCTCGAGACGCGAC	300			
Qy	301	GCC	ACCGCTGGG	CAAGGCGGAGAC	CTCTGTAGGCTTCTCCGAATCCCGTTCGACCTCCAGC	360			
Db	301	GCC	ACCGCTGGG	CAAGGCGGAGAC	CTCTGTAGGCTTCTCCGAATCCCGTTCGACCTCCAGC	360			
Qy	361	CGT	GAGCGCG	CGCGGCTTAC	CTGAGAGACTGTCAAGAAAGGAGATGGAGCGGGAC	420			
Db	361	CGT	GAGCGCG	CGCGGCTTAC	CTGAGAGACTGTCTAAGAAAGGAGATGGAGCGGGAC	420			
Qy	421	AGG	CGGATCGCG	GAACGCTTGG	CGCTTGGCGGGCTTTCGGGTTCTGGCCACCCCTTTT	480			
Db	421	AGG	CGGATCGCG	GAACGCTTGG	CGCTTGGCGGGCTTTCGGGTTCTGGCCACCCCTTTT	480			
Qy	481	CCCT	CGGCGATCG	CAAGCAGGCTT	CTTAAGTTCCTCGACGCTCTTGCGCCGGAAAACTC	540			
Db	481	CCCT	CGGCGATCG	CAAGCAGGCTT	CTTAAGTTCCTCGACGCTCTTGCGCCGGAAAACTC	540			
Qy	541	CGGG	AACCCCACT	GTCTTCTCTG	CCGACCGCAGACTCGGGTCAGTTACTGACGAA	600			
Db	541	CGGG	AACCCCACT	GTCTTCTCTG	CCGACCGCAGACTCGGGTCAGTTACTGACGAA	600			
Qy	601	ACT	GTCTCTCC	CAGCTCCTT	CGCCGCTCCCGGATTTGCTTCAAGAGGTGCTAAATTTGGAG	660			
Db	601	ACT	GTCTCTCC	CAGCTCCTT	CGCCGCTCCCGGATTTGCTTCAAGAGGTGCTAAATTTGGAG	660			
Qy	661	CCAA	CTTTTCGT	TGGAAATGTTT	CCGACGATGGCTAGATTTGCTGGAGTCTACAGCGC	720			
Db	661	CCAA	CTTTTCGT	TGGAAATGTTT	CCGACGATGGCTAGATTTGCTGGAGTCTACAGCGC	720			
Qy	721	CCT	GAGGCCCT	GAAGGAGCGG	GAACACAGCGCCGCCCCACAGCGCAGAAATCTTTGAG	780			
Db	721	CCT	GAGGCCCT	GAAGGAGCGG	GAACACAGCGCCGCCCCACAGCGCAGAAATCTTTGAG	780			

Qy	781	TT	CGCTG	CAGCTCG	ACTCCTC	CAGACCCCTCGGT	CACCACTCCCTTGATTGGCTAGAGGA	840																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																													
Db	781	TT	CGCTG	CAGCTCG	ACTCCTC	CAGACCCCTCGGT	CACCACTCCCTTGATTGGCTAGAGGA	840																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																													
Qy	841	GGG	GATCC	ACTTGG	CAATACT	TGCGCCCCCAGACCT	TAAATTTGGAGCTTAAAGGCCAAGGGAAG	900																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																													
Db	841	AGG	GATCC	ACTTGG	CAATACT	TGCGCCCCCAGACCT	TAAATTTGGAGCTTAAAGGCCAAGGGAAG	900																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																													
Qy	901	TG	CTTTT	GGACCC	TG	CAGCACAGGCTTTTCT	TTCTTAGAGCAGCAGCTGTGGGAGTGGAGCT	960																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																													
Db	901	TG	CTTTT	GGACCC	TG	CAGCACAGGCTTTTCT	TTCTTAGAGCAGCAGCTGTGGGAGTGGAGCT	960																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																													
Qy	961	GT	TGCC	CAGTAG	CTTCAAT	CCCGCTCTGTACT	CTTAACCGGGAACCTTGGCTCTTCGCCCTC	1020																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																													
Db	961	GT	TGCC	CAGTAG	CTTCAAT	CCCGCTCTGTACT	CTTAACCGGGAACCTTGGCTCTTCGCCCTC	1020																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																													
Qy	1021	TGG	GCCTCT	TAAACAT	TCAAC	GCATAGACAAATTT	CAGTGTGGTATCTCTATTTGCTGAAACCC	1080																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																													
Db	1021	TGG	GCCTCT	TAAACAT	TCAAC	GCATAGACAAATTT	CAGTGTGGTATCTCTATTTGCTGAAACCC	1080																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																													
Qy	1081	TT	CTTAC	TCTGGA	CTGCTT	TCTTAGGCT	TAGAGTCAGCTATCAGAAACAGTGTGGAATAG	1140																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																													
Db	1081	TT	CTTAC	TCTGGA	CTGCTT	TCTTAGGCT	TAGAGTCAGCTATCAGAAACAGTGTGGAATAG	1140																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																													
Qy	1141	CG	AGGT	AGT	CGGCTT	CC	AGACACTAACC	CAGAGACGCTGCTGAGAGAGGACCAATTG	1200																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																												
Db	1141	CG	AGGT	AGT	CGGCTT	CC	AGACACTAACC	CAGAGACGCTGCTGAGAGAGGACCAATTG	1200																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																												
Qy	1201	TC	AT	CCCCAG	CGCTG	AGTGCAGAACTCA	TTCCGGGCTCGTGGCAGGGATGTCCACCTCT	1260																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																													
Db	1201	TC	AT	CCCCAG	CGCTG	AGTGCAGAACTCA	TTCCGGGCTCGTGGCAGGGATGTCCACCTCT	1260																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																													
Qy	1261	TT	CT	ACG	AGGCGCT	TAC	CGAATTCACCATCTTCCCATGAACGCGCTGGAATTCCTTCA	1320																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																													
Db	1261	TT	CT	ACG	AGGCGCT	TAC	CGAATTCACCATCTTCCCATGAACGCGCTGGAATTCCTTCA	1320																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																													
Qy	1321	AC	AGGCT	TAA	CAAGGGG	CAAGATTT	TACCCACCCCTGAC	CAGGATAATGGCTACCAAGCCT	1380																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																												
Db	1321	AC	AGGCT	TAA	CAAGGGG	CAAGATTT	TACCCACCCCTGAC	CAGGATAATGGCTACCAAGCCT	1380																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																												
Qy	1381	GG	AGGAG	GAAC	CACAGCCTT	CTCGGAT	TGATCCAAACACT	TCGACAGATTAACCCACACA	1440																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																												
Db	1381	GG	AGGAG	GAAC	CACAGCCTT	CTCGGAT	TGATCCAAACACT	TCGACAGATTAACCCACACA	1440																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																												
Qy	1441	GT	TT	TGTTCT	GTCTGT	CGAGACAT	TTCTGAAAACACC	CAGGAATCCACTGAAGAAAAAT	1500																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																												
Db	1441	GT	TT	TGTTCT	GTCTGT	CGAGACAT	TTCTGAAAACACC	CAGGAATCCACTGAAGAAAAAT	1500																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																												
Qy	1501	AGA	ATT	TTA	ATCT	TACAGAGT	TCCATTGCTTTGG	AAGAAAGAGAGCCTTCTGAGGGCTG	1560																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																												
Db	1501	AGA	ATT	TTA	ATCT	TACAGAGT	TCCATTGCTTTGG	AAGAAAGAGAGAGCCTTCTGAGGGCTG	1560																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																												
Qy	1561	T	C	ACT	CTAGT	GAGATACCT	TATG	GAAGAGGACCTGGAGAGGCGCGAATAGTGTAGTTGA	1620																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																												
Db	1561	T	C	ACT	CTAGT	GAGATACCT	TATG	GAAGAGGACCTGGAGAGGCGCGAATAGTGTAGTTGA	1620																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																												
Qy	1621	T	T	ACT	CT	CAT	ACT	TAGAAGTGAACCTTCCCA	TTTCTG	CCAGACACAGCTTGTAGTAAACAACT	1680																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																										
Db	1621	T	T	ACT	CT	CAT	ACT	TAGAAGTGAACCTTCCCA	TTTCTG	CCAGACACAGCTTGTAGTAAACAACT	1680																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																										
Qy	1681	G	A	T	A	T	A	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T

Qy	1861	TCCTATAATATCCCCAGAACTTTTACAGCAACAATTCAGACTGCTGCGCAAGAAATTTGTTCTCTGA	1920
Db	1861	TCCTATAATATCCCCAGAACTTTTACAGCAACAATTCAGACTGCTGCGCAAGAAATTTGTTCTCTGA	1920
Qy	1921	AGAGCCCTTCGATTCACAGAGAGATTTGTCTGGCAAGTCTGATCTAGAGAAATTCCTCCCA	1980
Db	1921	AGAGCCCTTCGATTCACAGAGAGATTTGTCTGGCAAGTCTGATCTAGAGAAATTCCTCCCA	1980
Qy	1981	GTCTGGAAGCTTCTCTGAGACCCCTGAGCAGTAGTTCTGGGGAGGAAGATGACTGGGAATC	2040
Db	1981	GTCTGGAAGCTTCTCTGAGACCCCTGAGCAGTAGTTCTGGGGAGGAAGATGACTGGGAATC	2040
Qy	2041	TAGTCAGATGAAGCAGAGAGTCTCAAACTGTGGAACCTATCTGTATTTCTGATGACCC	2100
Db	2041	TAGTCAGATGAAGCAGAGAGTCTCAAACTGTGGAACCTATCTGTATTTCTGATGACCC	2100
Qy	2101	CTACAACTTTAAATTTTAAAGGCTCTCTTTTCAAACTATCAGGGGAAATGAGAAAGGCTG	2160
Db	2101	CTACAACTTTAAATTTTAAAGGCTCTCTTTTCAAACTATCAGGGGAAATGAGAAAGGCTG	2160
Qy	2161	TCGTGACTCAAAAGACCCCATCTGAGTCCATTTGTGGCCATTTCTGAGTGTCAACCTTACT	2220
Db	2161	TCGTGACTCAAAAGACCCCATCTGAGTCCATTTGTGGCCATTTCTGAGTGTCAACCTTACT	2220
Qy	2221	TTCTTTGAAAGTGCAGCTGTTGGGAGCCCAAGAAAGTGAATGTCCAGACTCGGTACAGCG	2280
Db	2221	TTCTTTGAAAGTGCAGCTGTTGGGAGCCCAAGAAAGTGAATGTCCAGACTCGGTACAGCG	2280
Qy	2281	TGAGTTCTTCTGGAGGAGACACACATGTCAAAAGAAAAGTAACTTCTCTTGA	2340
Db	2281	TGAGTTCTTCTGGAGGAGACACACATGTCAAAAGAAAAGTAACTTCTCTTGA	2340
Qy	2341	AGAAGTTACTGAGTATATATAAGTGTGTAGAGATCGCAAGGACCATGGGAAGATTT	2400
Db	2341	AGAAGTTACTGAGTATATATAAGTGTGTAGAGATCGCAAGGACCATGGGAAGATTT	2400
Qy	2401	TGCAAGGATGATGTCAGGTTCCAGAAACGAATTCAGAAACAGAAAGATCTATTGATA	2460
Db	2401	TGCAAGGATGATGTCAGGTTCCAGAAACGAATTCAGAAACAGAAAGATCTATTGATA	2460
Qy	2461	TTGCTTGACATTTGAACACAGAGAGAGATTTTAAATAGACTCCAGGGAACATGCTTCAA	2520
Db	2461	TTGCTTGACATTTGAACACAGAGAGAGATTTTAAATAGACTCCAGGGAACATGCTTCAA	2520
Qy	2521	AGGACTTAATGCTCTCAAGCAATGTGAGTTGGCAGCTGTAGTCTAGCTAGCATACAC	2580
Db	2521	AGGACTTAATGCTCTCAAGCAATGTGAGTTGGCAGCTGTAGTCTAGCATACAC	2580
Qy	2581	TACCTCTTACCTGAGAGGTGCTTTTAAACAACTTTGGCAGCTGCTCTTGGACATTT	2640
Db	2581	TACCTCTTACCTGAGAGGTGCTTTTAAACAACTTTGGCAGCTGCTCTTGGACATTT	2640
Qy	2641	TTTTTTTTTAGAGGAATGTAATCTGGATCTAGTTTAAATTTTTTTTTTTTGGCAATATCCC	2700
Db	2641	TTTTTTTTTAGAGGAATGTAATCTGGATCTAGTTTAAATTTTTTTTTTTTGGCAATATCCC	2700
Qy	2701	ACTCAGAAACATTCAGGTTTCAAGCCGCTGTATGAAGAGTGAAGTGTGATTT	2760
Db	2701	ACTCAGAAACATTCAGGTTTCAAGCCGCTGTATGAAGAGTGAAGTGTGATTT	2760
Qy	2761	CTAATCTCTCCCTTTTGAATTTAGTTGAGATGCTTTTTTAAATGTCTTGCCTGCATGAG	2820
Db	2761	CTAATCTCTCCCTTTTGAATTTAGTTGAGATGCTTTTTTAAATGTCTTGCCTGCATGAG	2820
Qy	2821	GTGGAAGGGACCTTTTGTAGTTGTCTATTTTGCATTTTCAAACTTATTTCTTGGAAA	2880
Db	2821	GTGGAAGGGACCTTTTGTAGTTGTCTATTTTGCATTTTCAAACTTATTTCTTGGAAA	2880
Qy	2881	ACAATATTTATAGGCTTAAAGCCCATTTTCAATTTCTAATCTAATATGTCCTATC	2940
Db	2881	ACAATATTTATAGGCTTAAAGCCCATTTTCAATTTCTAATCTAATATGTCCTATC	2940
Qy	2941	TG 2942	

Db	2941	TG 2942	
RESULT 4			
LOCUS	BC065280		
DEFINITION	Homo sapiens cDNA clone MGC:74824 IMAGE:6172811, complete cds.		
ACCESSION	BC065280		
VERSION	BC065280.1	GI:40807094	
KEYWORDS	MGC.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
REFERENCE	1 (bases 1 to 5264)		
AUTHORS	Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zebberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Uedin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullaly, S.J., Bonak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smallos, D.E., Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.		
TITLE	Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences		
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)		
PUBMED	12477932		
REFERENCE	2 (bases 1 to 5264)		
AUTHORS	Director MGC Project.		
TITLE	Direct Submission		
JOURNAL	Submitted (12-JAN-2004) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA		
REMARK	NIH-MGC Project URL: <a href="http://mgc.nci.nih.gov">http://mgc.nci.nih.gov</a>		
COMMENT	Contact: MGC help desk Email: <a href="mailto:cgabs-x@mail.nih.gov">cgabs-x@mail.nih.gov</a> Tissue Procurement: AFCC cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC), Gaithersburg, Maryland; Web site: <a href="http://www.nisc.nih.gov/">http://www.nisc.nih.gov/</a> Contact: <a href="mailto:nisc_mgc@nih.gov">nisc_mgc@nih.gov</a> Akhter, N., Ayale, K., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W., Bouffard, G.G., Green, K., Brinkley, C., Brooks, S., Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P., Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Laric, P., Legaspi, R., Maduro, Q.L., Masiello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Stantripop, S., Thomas, P.J., Touchman, J.W., Taurgeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L., Young, A., Zhang, L.-H. and Green, E.D.		
FEATURES	Location/Qualifiers		
source	1. 5264 /organism="Homo sapiens"		

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAC Plate: 140 Row: p Column: 3  
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 41282220.

```
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:74824 IMAGE:6172811"
/tissue_type="Eye, retinoblastoma"
/clone_lib="NIH_MGC_67"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
381..2522
CDS
/codon_start=1
/product="Unknown (protein for MGC:74824)"
/protein_id="AAH65280.1"
/db_xref="GI:40807095"
/translation="MEPTGGSRRKLRPRGFRFWPPFRRSQSSKFFPTLPGEN
SNPTLLSSAOPETKRVSYWTKLLSOLLAPLGLKLOKLIWSQLFGMPPTRWLDFAGV
YSALRALGRKREPAAPTAKLSLSIQLSDSPVTSPLDMLBEGIHMOYSPPDLKLEL
KAKGSAIDPAQAQFLEQLQWGVLLPSSLASRLYSNRELSSPSGLPINIORDNFVS
VSYLNPFLDYOCFPLEVSYQNSDGNSEVGFQITLPRESSCIREDHCHPQPSAELIP
ASWQCCPLSTEGLEPIIHLHMKRLLEFLQAKGQDLPTPDQNGYSLHEEHLIRM
DPKHCRDNPOTFPAAGDI PGNTQSTPEKEI ELITTEVPLALEESSESGPSESIEM
EKEPGEGRISVVDYSYLEDLP ISARPACSNKLIDYILGASLSLETSDPDEGDIWE
RAEDDGFDSLSLSDLEODPEGLHANSFCSVDPNPQNFATIoTAAIRIVPEEPS
DSEKDLGKSLDENSQSGLPETPEHSSGREDDWESSADEASLKLWNSFCNSDDPY
NPLNKAFFQTSGENEKCRDSETPSEISALISECHTLLSKVQLLGSQSECEPDSVQ
RDVLSGRHHVKKRKKVFLBEVTEYIISGDEDRKPWEEFARDGCRFQKRQIETEDA
IGYCLTFEHRERMFNRLOGTCTFKGLNVLKQC"
```

ORIGIN

Query Match		99.0%; Score 2912.8; DB 9; Length 5264;
Best Local Similarity		99.9%; Pred. No. 0;
Matches 2914; Conservative 0; Mismatches 2; Indels 0; Gaps 0;		
QY	27	AGCCGGCCCTACCCAGTCCCTCCGGTATCGCGTGTGCTCAGGGGCTTTTCAACCCCTCTGTCA 86
DB	1	AGCCGGCCCTACCCAGTCCCTCCGGTATCGCGTGTGCTCAGGGGCTTTTCAACCCCTCTGTCA 60
QY	87	GTGGAACCAATCGCCGAGGCCGTGGGGGACCTCTATCCATGATGTTGAAGCGTGGAG 146
DB	61	GTGGAACCAATCGCCGAGGCCGTGGGGGACCTCTATCCATGATGTTGAAGCGTGGAG 120
QY	147	CGGACTAGGAACCTCTTCCCGCCGAGTGAAGTGCATCGTCCGCCCTTATTCGG 206
DB	121	CGGACTAGGAACCTCTTCCCGCCGAGTGAAGTGCATCGTCCGCCCTTATTCGG 180
QY	207	CGGGCTGTTCTTCCCTGTGTTCTGCGCCCGTCCGCAATCGCTGCCCTCTGTGGCTTT 266
DB	181	CGGGCTGTTCTTCCCTGTGTTCTGCGCCCGTCCGCAATCGCTGCCCTCTGTGGCTTT 240
QY	267	TCTGCTGGTTCGAAGATCGGCTGAGCAGCGAGCGCCACCGCTGGGCAAGGCCGAGACTC 326
DB	241	TCTGCTGGTTCGAAGATCGGCTGAGCAGCGAGCGCCACCGCTGGGCAAGGCCGAGACTC 300
QY	327	TGTAGGCTTCCTCCGAATCCGTCGACCTCCAGCCGCTGAGCGCGCGGCCCTACCTGAG 386
DB	301	TGTAGGCTTCCTCCGAATCCGTCGACCTCCAGCCGCTGAGCGCGCGGCCCTACCTGAG 360
QY	387	AGACTGTCAAGAAAAGGAGATGGAGCCGGGACAGCGGGATCGCGGAAAACGGCTTGGCC 446
DB	361	AGACTGTCAAGAAAAGGAGATGGAGCCGGGACAGCGGGATCGCGGAAAACGGCTTGGCC 420
QY	447	CTCGGGCGGGCTTCGGGTTCTGGCCACCGCTTTTTCCTTCGGGATTCGCAAGAGGCTTT 506
DB	421	CTCGGGCGGGCTTCGGGTTCTGGCCACCGCTTTTTCCTTCGGGATTCGCAAGAGGCTTT 480
QY	507	CTAAGTTCGAGCGCTCTTGGCCCGGAAACTCCGGGAAACCCACACTGTTTCTCTCTG 566
DB	481	CTAAGTTCGAGCGCTCTTGGCCCGGAAACTCCGGGAAACCCACACTGTTTCTCTCTG 540
QY	567	CCCAGCCGAGACTCGGGTCAGTACTGGACGAAACTGCTCTCCAGCTCTCTTGGCGGC 626
DB	541	CCCAGCCGAGACTCGGGTCAGTACTGGACGAAACTGCTCTCCAGCTCTCTTGGCGGC 600
QY	627	TCCCGGATTTGTTTCAAGAGGTGCTTAATTTGGAGCCAACTTTTCGGTGAATGTTCCGA 686
DB		

DB	601	TCCCGGATTGCTTTCAAGAGGTGCTAAATTTGGAGCCAACTTTTCGTGGGAATGTTCCGA 660
QY	687	CCAGATGCTAGATTTTCTGGAGTCTACAGCGCCCTGAGAGCCCTGAGAGCGAGCGGAGA 746
DB	661	CCAGATGCTAGATTTTCTGGAGTCTACAGCGCCCTGAGAGCCCTGAGAGCGAGCGGAGA 720
QY	747	AACCAGCGCGCCCCACAGCGCAGAAATCTTTTGGAGTTCGCTGCACTCGACTCTCTCAGACC 806
DB	721	AACCAGCGCGCCCCACAGCGCAGAAATCTTTTGGAGTTCGCTGCACTCGACTCTCTCAGACC 780
QY	807	CCTCGGTACAGTCCCTCTGATTTGGCTAGAGGAGGGATCCACTGCAATACTTCGCCCC 866
DB	781	CCTCGGTACAGTCCCTCTGATTTGGCTAGAGGAGGGATCCACTGCAATACTTCGCCCC 840
QY	867	CAGACCTAAAATTTGGAGCTTAAAGCCCAAGGAAGTCTTTGGACCTCTGCAAGCAGAGCTT 926
DB	841	CAGACCTAAAATTTGGAGCTTAAAGCCCAAGGAAGTCTTTTGGACCTCTGCAAGCAGAGCTT 900
QY	927	TTCTCTTTAGAGCAGCAGCTGTGGGGAGTGGAGCTGTTTCCCAGTAGCTTTCAATCCCGTC 986
DB	901	TTCTCTTTAGAGCAGCAGCTGTGGGGAGTGGAGCTGTTTCCCAGTAGCTTTCAATCCCGTC 960
QY	987	TGTACTCTAAACCGGAACTTTGGCTTCTCGCCCTCTGGGCTCTTAAACATTTCAACGCAATAG 1046
DB	961	TGTACTCTAAACCGGAACTTTGGCTTCTCGCCCTCTGGGCTCTTAAACATTTCAACGCAATAG 1020
QY	1047	ACAAATTTCAAGTGTGATCTTATTTGCTGAACCTCTCTACCTGGAGTGTCTTCTTAGGC 1106
DB	1021	ACAAATTTCAAGTGTGATCTTATTTGCTGAACCTCTCTACCTGGAGTGTCTTCTTAGGC 1080
QY	1107	TAGAAGTCAGCTATTCAGAACAGTGAATGGAATAGCAGGAGTAGTGGCTTCCAGACACTAA 1166
DB	1081	TAGAAGTCAGCTATTCAGAACAGTGAATGGAATAGCAGGAGTAGTGGCTTCCAGACACTAA 1140
QY	1167	CCCAGAGAGAGCTGCTGAGAGAGGACCATTTGTCTATCCCAGCCGCTGAGTGCAGAAC 1226
DB	1141	CCCAGAGAGAGCTGCTGAGAGAGGACCATTTGTCTATCCCAGCCGCTGAGTGCAGAAC 1200
QY	1227	TCATTCCGGCTGTGGCAGGGATGCCACTCTTTCTACGGAAGGCTTACCAGAAATTC 1286
DB	1201	TCATTCCGGCTGTGGCAGGGATGTCCACTCTTTCTACGGAAGGCTTACCAGAAATTC 1260
QY	1287	ACCATTTTCATGAAACCGCTGGAATTCCTTCAACAGGCTAAACAGGGGCAAGATTAC 1346
DB	1261	ACCATTTTCATGAAACCGCTGGAATTCCTTCAACAGGCTAGCAAGGGCAAGATTAC 1320
QY	1347	CCACCCCTGACAGGTAATGGCTACCAAGCCCTGGAGAGGACACAGCCTTCTCCGGA 1406
DB	1321	CCACCCCTGACAGGTAATGGCTACCAAGCCCTGGAGAGGAAACACAGCCTTCTCCGGA 1380
QY	1407	TGGATCCAAAACACTGCGAGAGATAAACCCAAACACAGTTTTGTCTGCTGCTGGAGACATTC 1466
DB	1381	TGGATCCAAAACACTGCGAGAGATAAACCCAAACACAGTTTTGTCTGCTGCTGGAGACATTC 1440
QY	1467	CTGGAACAACCCAGGAATCCACTGAAGAAAAATAGAAATTTAATCTACAGAGGTTCCAC 1526
DB	1441	CTGGAACAACCCAGGAATCCACTGAAGAAAAATAGAAATTTAATCTACAGAGGTTCCAC 1500
QY	1527	TTGCTTTTGAAGAAGAGAGCCCTTCTGAGGCTGTCCATCTAGTGAGATACCTATGGA 1586
DB	1501	TTGCTTTTGAAGAAGAGAGCCCTTCTGAGGCTGTCCATCTAGTGAGATACCTATGGA 1560
QY	1587	AGGAGCTCGAGAGGGCCGAATAGTGTAGTTGATTAATCTCATCTAGAGAGGTGACCTTC 1646
DB	1561	AGGAGCTCGAGAGGGCCGAATAGTGTAGTTGATTAATCTCATCTAGAGAGGTGACCTTC 1620
QY	1647	CCATTTCTGCAGACCGAGCTTGTAGTAAACAACTGATAGATTATTTTGGAGGTTGCAT 1706
DB	1621	CCATTTCTGCAGACCGAGCTTGTAGTAAACAACTGATAGATTATTTTGGAGGTTGCAT 1680
QY	1707	CCAGTGAACCTGGAACCAAGTTCCTGATCCAGAGGTGAGGATTTGGATGAGGAACTCAGG 1766
DB	1681	CCAGTGAACCTGGAACCAAGTTCCTGATCCAGAGGTGAGGATTTGGATGAGGAACTCAGG 1740





		Matches 2332;	Conservative	0;	Mismatches	5;	Indels	0;	Gaps	0;
Qy	1	ATTTTGGGCTTCGCTTCCACCGCACCGCGCCCTACCGTCCCTTCGGGTATCGCGTTG	60							
Db	63775	ATTTTGGGCTTCGCTTCCACCGCACCGCGCCCTACCGTCCCTTCGGGTATCGCGTTG	63716							
Qy	61	CTCAGGGCTTTTCAACCCCTCTGTGTCAGTTCGGAACCATCGCCGAGGCGCTGGGGGACT	120							
Db	63715	CTCAGGGCTTTTCAACCCCTCTGTGTCAGTTCGGAACCATCGCCGAGGCGCTGGGGGACT	63656							
Qy	121	CCTATCCATGGTGTGAAGCGTCGAGCCGACCTAGGGAACTCTCTCCCGCCGAGGATGGA	180							
Db	63655	CCTATCCATGGTGTGAAGCGTCGAGCCGACCTAGGGAACTCTCTCTCCCGCCGAGGATGGA	63596							
Qy	101	AGTCGCATCAGTCGCGCGCTATTGCGCGGCTGTCTTCCTCTGTCTGCGCGCCGCTG	240							
Db	63595	AGTCGCATCAGTCGCGCGCTATTGCGCGGCTGTCTTCCTCTGTCTGCGCGCCGCTG	63536							
Qy	241	CGCGATTCCGCTGCTCTGTGCTTTTCTGCTGGCTCGAAGATCGGCTCGAGCAGCGAC	300							
Db	63535	CGCGATTCCGCTGCTCTGTGCTTTTCTGCTGGCTCGAAGATCGGCTCGAGCAGCGAC	63476							
Qy	301	GCCACCGCTGGGCAAGGCGGAGACTCTGTAGGCTTCTTCGAACTCCCGTCCAGCTCCAGC	360							
Db	63475	GCCACCGCTGGGCAAGGCGGAGACTCTGTAGGCTTCTTCGAACTCCCGTCCAGCTCCAGC	63416							
Qy	361	CGCTGAGCGCGCGCCCTACTCTGAGAGACTGTCAAGAAAGAGATGGAGCGGGGAC	420							
Db	63415	CGCTGAGCGCGCGCCCTACTCTGAGAGACTGTCAAGAAAGAGATGGAGCGGGGAC	63356							
Qy	421	AGGCGGATCGCGGAAACGGCTTGGCCCTCGGCGGGCTTTCGGTCTTGCCCAACCTTTT	480							
Db	63355	AGGCGGATCGCGGAAACGGCTTGGCCCTCGGCGGGCTTTCGGTCTTGCCCAACCTTTT	63296							
Qy	481	CCCTCGGCGATCGCAAGCAGGCTCTTCTAAGTTCGCGAGCGCTCTTGCCCGGAAAATC	540							
Db	63295	CCCTCGGCGATCGCAAGCAGGCTCTTCTAAGTTCGCGAGCGCTCTTGCCCGGAAAATC	63236							
Qy	541	CGGGAAACCCACACTGCTTCTCTGCGCCAGCCGAGACTCGGGTCAGTCTACTTGACGAA	600							
Db	63235	CGGGAAACCCACACTGCTTCTCTGCGCCAGCCGAGACTCGGGTCAGTCTACTTGACGAA	63176							
Qy	601	ACTGCTCTCCAGCTCTTGGCGCGCTCCCGGATTTGCTTCAGAAAGTGTCTAAATTTGGAG	660							
Db	63175	ACTGCTCTCCAGCTCTTGGCGCGCTCCCGGATTTGCTTCAGAAAGTGTCTAAATTTGGAG	63116							
Qy	661	CCAACTTTTCGGTGGAAATGTTTCCAGCAGATGGCTAGATTTTGTGCTGAGTCTACAGCGC	720							
Db	63115	CCAACTTTTCGGTGGAAATGTTTCCAGCAGATGGCTAGATTTTGTGCTGAGTCTACAGCGC	63056							
Qy	721	CTGTAGAGCCCTGAAGGGACGGGAGAAACCAAGCCGCCCCACAGCGCAGAAATCTTTGAG	780							
Db	63055	CTGTAGAGCCCTGAAGGGACGGGAGAAACCAAGCCGCCCCACAGCGCAGAAATCTTTGAG	62996							
Qy	781	TTGCGTCGAGCTCGACTCTCTCAGACCCCTCGGTCAACAGTCCCTTGTGATGGCTAGAGGA	840							
Db	62995	TTGCGTCGAGCTCGACTCTCTCAGACCCCTCGGTCAACAGTCCCTTGTGATGGCTAGAGGA	62936							
Qy	841	GGGGATCACTGGGCAATFACTCGCCGCCAGACTTAAATTTGGAGCTTAAGCCGAGGGAAG	900							
Db	62935	GGGGATCACTGGGCAATFACTCGCCGCCAGACTTAAATTTGGAGCTTAAGCCGAGGGAAG	62876							
Qy	901	TGCTTTGGAACCTCGACGACAGGCTTTTCTCTTAGAGCAGCAGCTGTGGGAGTGGAGCT	960							
Db	62875	TGCTTTGGAACCTCGACGACAGGCTTTTCTCTTAGAGCAGCAGCTGTGGGAGTGGAGCT	62816							
Qy	961	GTTCGCCAGTACCTTCGAATCCCGTCTGTACTCTTAACCCGGGAACTTGGCTCTTCGCCCTC	1020							
Db	62815	GTTCGCCAGTACCTTCGAATCCCGTCTGTACTCTTAACCCGGGAACTTGGCTCTTCGCCCTC	62756							
Qy	1021	TGGGCTCTTAACATTTCAACGATAGACAAATTTCAAGTGTGGTATCCTATTTCGTGAACCC	1080							
Db	62755	TGGGCTCTTAACATTTCAACGATAGACAAATTTTCAGTGTGGTATCCTATTTCGTGAACCC	62696							

Qy	1081	TTCTTACTGGAAGTCTTCTTAGGCTAGAACTCAGTATCAGAACTAGTGTGAAATAG	1140
Db	62695	TTCTTACTGGAAGTCTTCTTAGGCTAGAACTCAGTATCAGAACTAGTGTGAAATAG	62636
Qy	1141	CGAGGTAGTGGCTTCAGACACTAACCCAGAGAGAGCTGCTTGAGAGAGGACCAATTG	1200
Db	62635	CGAGGTAGTGGCTTCAGACACTAACCCAGAGAGAGCTGCTTGAGAGAGGACCAATTG	62576
Qy	1201	TCATCCCCAGCGCTGAGTCAGAACTCATTTCCGGCTCTGGGAGGAGTGTCCACCTCT	1260
Db	62575	TCATCCCCAGCGCTGAGTCAGAACTCATTTCCGGCTCTGGGAGGAGTGTCCACCTCT	62516
Qy	1261	TTCTACGGAAGGCTACCCAGAAATTCACATCTTCCGATGAAACCGCTGGAAATTCCTTCA	1320
Db	62515	TTCTACGGAAGGCTACCCAGAAATTCACATCTTCCGATGAAACCGCTGGAAATTCCTTCA	62456
Qy	1321	ACAGGCTAAACAGGGGCAAGATTTACCCACCCCTGACAGGATTAATGGCTACCAAGCT	1380
Db	62455	ACAGGCTAAACAGGGGCAAGATTTACCCACCCCTGACAGGATTAATGGCTACCAAGCT	62396
Qy	1381	GGAGGAGAAACACAGCGCTTCTCCGGATGGATCCAAAACACTGCAGAGATAACCCAAACA	1440
Db	62395	GGAGGAGAAACACAGCGCTTCTCCGGATGGATCCAAAACACTGCAGAGATAACCCAAACA	62336
Qy	1441	GTCTTGTCTCTGCTGGAGACATTCCTTGAAACACCCAGGAATCCACTGAAGAAAAAAT	1500
Db	62335	GTCTTGTCTCTGCTGGAGACATTCCTTGAAACACCCAGGAATCCACTGAAGAAAAAAT	62276
Qy	1501	AGAAATTAATTAACATACAGAGTTCCACTTGTCTTGGAGAGAGAGCCCTTCTGAGGCTG	1560
Db	62275	AGAAATTAATTAACATACAGAGTTCCACTTGTCTTGGAGAGAGAGCCCTTCTGAGGCTG	62216
Qy	1561	TCCATCTAGTGGAGATCCTATGAAAAGAGAGCCCTGGAGAGGCGCAATAAGTGTAGTTGA	1620
Db	62215	TCCATCTAGTGGAGATCCTATGAAAAGAGAGCCCTGGAGAGGCGCAATAAGTGTAGTTGA	62156
Qy	1621	TTACTCATACCTAGAAGGTGACCTTCCCAATTTCTGCAGACCCAGCTTGTAGTAACAACT	1680
Db	62155	TTACTCATACCTAGAAGGTGACCTTCCCAATTTCTGCAGACCCAGCTTGTAGTAACAACT	62096
Qy	1681	GATGATATTAATTTTGGAGGTGATCCAGTGCCTTGGAAACAGTTCGTATCCAGAGG	1740
Db	62095	GATGATATTAATTTTGGAGGTGATCCAGTGCCTTGGAAACAGTTCGTATCCAGAGG	62036
Qy	1741	TGAGATTTGGATGAGGAAGCTGAGGATGATGGTTTTGTAGTATAGTCACTGTGAGA	1800
Db	62035	TGAGATTTGGATGAGGAAGCTGAGGATGATGGTTTTGTAGTATAGTCACTGTGAGA	61976
Qy	1801	CTCAGACCTTTGAACAAGACCCCTGAAGGGCTTACCTTTTGGAACTCTTCTGCACTGTAGA	1860
Db	61975	CTCAGACCTTTGAACAAGACCCCTGAAGGGCTTACCTTTTGGAACTCTTCTGCACTGTAGA	61916
Qy	1861	TCCTTATTAATCCCGCAAGACTTTACAGCAACAAATTCAGACTGCTGCCAGNAATGTTCTGA	1920
Db	61915	TCCTTATTAATCCCGCAAGACTTTACAGCAACAAATTCAGACTGCTGCCAGNAATGTTCTGA	61856
Qy	1921	AGAGCTTTCTGATTCAGAGGAAGGATTTGTCTGGCAAGTCTGATCTAGAGAAATTCCTCCA	1980
Db	61855	AGAGCTTTCTGATTCAGAGGAAGGATTTGTCTGGCAAGTCTGATCTAGAGAAATTCCTCCA	61796
Qy	1981	GTCTGGAAGCTTCTCTGAGACCCCTGAGCATPAGTTCCTGGGGAGGAAGATGATCGGGAATC	2040
Db	61795	GTCTGGAAGCTTCTCTGAGACCCCTGAGCATPAGTTCCTGGGGAGGAAGATGATCGGGAATC	61736
Qy	2041	TAGTGCAGATGAAAGCAGAGGCTCTCAAACTGTGAACTCAATTCGTGAATTCGTATGACCC	2100
Db	61735	TAGTGCAGATGAAAGCAGAGGCTCTCAAACTGTGAACTCAATTCGTGAATTCGTATGACCC	61676
Qy	2101	CTACAAACCTTTAAATTTTAAAGGCTCTCTTTTCAAACTACGGGGAATAAGAAAGGCTG	2160
Db	61675	CTACAAACCTTTAAATTTTAAAGGCTCTCTTTTCAAACTACGGGGAATAAGAAAGGCTG	61616

```
Qy 2161 TCGTACTCAAGACCCCATCTGAGTCATTTGGGCAATTTCTGAGTGTACACCTTACT 2220
Db 61615 TCGTACTCAAGACCCCATCTGAGTCATTTGGGCAATTTCTGAGTGTACACCTTACT 61556
Qy 2221 TTCTTTGAAGTGCAGCTGTGGGAGCCAAAGAAAGTGAATGTCCACACTCGGTACAGCG 2280
Db 61555 TTCTTTGAAGTGCAGCTGTGGGAGCCAAAGAAAGTGAATGTCCACACTCGGTACAGCG 61496
Qy 2281 TGAGCTTCTTTCTTGAGGAAGACACACATGTCAAAAGAAAAAAGGTAACTTTCCT 2337
Db 61495 TGAGCTTCTTTCTTGAGGAAGACACACATGTCAAAAGAAAAAAGGTGTGTTCCT 61439

RESULT 6
AL627202/c 328863 bp DNA linear HTG 24-OCT-2002
LOCUS Homo sapiens chromosome 1 clone RP11-174P23, 2 unordered pieces.
DEFINITION AL627202
ACCESSION AL627202
VERSION 6
KEYWORDS HTG; HTGS PHASE1; HTGS_CANCELLED.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Clark, S.
Direct Submission
Submitted (21-OCT-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk
On Oct 25, 2002 this sequence version replaced gi:20268846.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA174P23
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 170058 bases at least Q40
Consensus quality: 171825 bases at least Q30
Consensus quality: 172641 bases at least Q20
Insert size: 328763; sum-of-contigs
Insert size: 172230; 12.6% error; agarose-fp
Quality coverage: 2.67x in Q20 bases; sum-of-contigs Quality
coverage: 5.15x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 6943: contig of 5943 bp in length
* 6944 7043: gap of 100 bp
* 7044 328863: contig of 321820 bp in length.

FEATURES
source
    1..328863
        /organism="Homo sapiens"
        /mol_type="genomic DNA"
        /db_xref="taxon:9606"
        /chromosome="1"
        /clone="RP11-174P23"
        /clone_lib="RPCI-11.1"
        1..6943
        /note="assembly_fragment:03428"
        7044..328863
        /note="assembly_fragment:03558"

misc_feature
misc_feature
```

## ORIGIN

```
Query Match 79.2%; Score 2329; DB 2; Length 328863;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2332; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 ATTTTGGGCTTGGCTTCCACCGCACAGCGGGCTACCCAGTCTCTCCGGTATCGGTTG 60
Db 225562 ATTTTGGGCTTGGCTTCCACCGCACAGCGGGCTACCCAGTCTCTCCGGTATCGGTTG 225503
Qy 61 CTCAGGGGCTTTTCAACCCCTCTGTAGTCGGAACCATCGCCGAGCGCGTGGGGGACT 120
Db 225502 CTCAGGGGCTTTTCAACCCCTCTGTAGTCGGAACCATCGCCGAGCGCGTGGGGGACT 225443
Qy 121 CCTATCCATGGTGTGTAAGCGTCGAGCCGACCTAGGGAACCTCTCTCCCGCCAGGATGGA 180
Db 225442 CCTATCCATGGTGTGTAAGCGTCGAGCCGACCTAGGGAACCTCTCTCCCGCCAGGATGGA 225383
Qy 181 AGTCGATCATGTCGCCGCTATTTGCGGGGCTGTCTTCCGTGTGTCTGCGGCCGCTG 240
Db 225382 AGTCGATCATGTCGCCGCTATTTGCGGGGCTGTCTTCCGTGTGTCTGCGGCCGCTG 225323
Qy 241 CGSCATTCGCTGCCCTCTGTGGCTTTTCTGCTGGCTCGAAGATCGGCCCTGGAGCAGC 300
Db 225322 CGSCATTCGCTGCCCTCTGTGGCTTTTCTGCTGGCTCGAAGATCGGCCCTGGAGCAGC 225263
Qy 301 GCCACCGCTGGGCAAGCGCGGACACTCTGTAGGCTTCTCTCGAATCCCGTCCAGCTCCAGC 360
Db 225262 GCCACCGCTGGGCAAGCGCGGACACTCTGTAGGCTTCTCTCGAATCCCGTCCAGCTCCAGC 225203
Qy 361 CGCTAGCGCGCGGCTTACTCTGAGAGACTGTCAAGAAAAAGAGAGATGAGCGGGGAC 420
Db 225202 CGCTAGCGCGCGGCTTACTCTGAGAGACTGTCAAGAAAAAGAGAGATGAGCGGGGAC 225143
Qy 421 AGCGGATCGCGGAACGCTTGGCCCTCGGGGGGCTTCCGGTCTTGGCCACCTTTT 480
Db 225142 AGCGGATCGCGGAACGCTTGGCCCTCGGGGGGCTTCCGGTCTTGGCCACCTTTT 225083
Qy 481 CCCTCGCGGATCGCAAGCAGGCTCTTCTAAGTTCCGAGCGCTCTTTGGCCCGGAAATC 540
Db 225082 CCCTCGCGGATCGCAAGCAGGCTCTTCTAAGTTCCGAGCGCTCTTTGGCCCGGAAATC 225023
Qy 541 CGGGAACCCGACACTCTCTTCTGCGCAGCCCGAGACTCGGTCAGTTACTTGAGCGAA 600
Db 225022 CGGGAACCCGACACTCTCTTCTGCGCAGCCCGAGACTCGGTCAGTTACTTGAGCGAA 224963
Qy 601 ACTGCTCTCCAGCTCTTGGCGCTCCCGGATGCTTCAAGAGTCTTAATTTGGAG 660
Db 224962 ACTGCTCTCCAGCTCTTGGCGCTCCCGGATGCTTCAAGAGTCTTAATTTGGAG 224903
Qy 661 CCAACTTTTTCGGTGGAAATGTTTCCGACCAAGATGGCTAGATTTTGTCTGAGTCTACAGCG 720
Db 224902 CCAACTTTTTCGGTGGAAATGTTTCCGACCAAGATGGCTAGATTTTGTCTGAGTCTACAGCG 224843
Qy 721 CTGTAGAGCCCTGAAGGAGCGGAGAAACAGCGCCGCCACACAGCGAGAAATCTTTGAG 780
Db 224842 CTGTAGAGCCCTGAAGGAGCGGAGAAACAGCGCCGCCACACAGCGAGAAATCTTTGAG 224783
Qy 781 TTGCTGCGAGCTCGACTCTCTCAGACCCCTCGGTCACAGTCCCTTTGATTGGCTAGAGGA 840
Db 224782 TTGCTGCGAGCTCGACTCTCTCAGACCCCTCGGTCACAGTCCCTTTGATTGGCTAGAGGA 224723
Qy 841 GGGGATCCCACTGGCAATATCTCGCCCCCAGACCTTAATTAATTTGAGGCTTAAGCCAAAGGAAG 900
Db 224722 GGGGATCCCACTGGCAATATCTCGCCCCCAGACCTTAATTAATTTGAGGCTTAAGCCAAAGGAAG 224663
Qy 901 TGCTTTGAGCCCTGAGAGCAGGCTTTTCTTTAGAGCAGCAGCTGTGGGAGTGGAGCT 960
Db 224662 TGCTTTGAGCCCTGAGAGCAGGCTTTTCTTTAGAGCAGCAGCTGTGGGAGTGGAGCT 224603
Qy 961 GTTGGCCAGTAGGCTTCAATCCGCTCTGACTCTTAACCGGAGTGGCTCTTTCGCCCTC 1020
Db 224602 GTTGGCCAGTAGGCTTCAATCCGCTCTGACTCTTAACCGGAGTGGCTCTTTCGCCCTC 224543
```

Qy	1021	TGGGCTCTAAACATTCAACGCATAGACAAATTTCAGTGTGGTATCCTATTTTCTGCTGAACCC	1080
Db	224542	TGGGCTCTTAAACATTTCAACGCATAGACAAATTTCAGTGTGGTATCCTATTTTCTGCTGAACCC	224483
Qy	1081	TTCTTACCTGGAGTCGTTTCTTAGGCTAGAACTCAGCTATCAGAACAGTGATGGAAATAG	1140
Db	224482	TTCTTACCTGGAGTCGTTTCTTAGGCTAGAACTCAGCTATCAGAACAGTGATGGAAATAG	224423
Qy	1141	CGAGGTAGTTCGGCTTCCAGACACTAACCCACAGAGCAGCTCCCTCTGAGAGAGCACCATG	1200
Db	224422	CGAGGTAGTTCGGCTTCCAGACACTAACCCACAGAGCAGCTCCCTCTGAGAGAGCACCATG	224363
Qy	1201	TCATCCCAGCGCTGAGTGCAGAACTCAATTCGGGCTCGTGGCAGGGGATGTCACCTCT	1260
Db	224362	TCATCCCAGCGCTGAGTGCAGAACTCAATTCGGGCTCGTGGCAGGGGATGTCACCTCT	224303
Qy	1261	TTCTACGGAAGCCCTACAGAAATTCACCATCTTCGCATGAACCGCTGGAAATTCCTTCA	1320
Db	224302	TTCTACGGAAGCCCTACAGAAATTCACCATCTTCGCATGAACCGCTGGAAATTCCTTCA	224243
Qy	1321	ACAGGCTAACAGGGGCAAGATTATCACCAACCCTGACACAGGATATGGCTACCAAGCCT	1380
Db	224242	ACAGGCTAACAGGGGCAAGATTATCACCAACCCTGACACAGGATATGGCTACCAAGCCT	224183
Qy	1381	GGAGGAGAAACAAGCCTTCTCCGATGGATCCAAAACACTGCAGAGATAACCCCAACACA	1440
Db	224182	GGAGGAGAAACAAGCCTTCTCCGATGGATCCAAAACACTGCAGAGATAACCCCAACACA	224123
Qy	1441	GTTTGTTCCTGCTCGTGGAGACATTCCTTGGAACACCCAGGAATCCACTGGAAGAAAAT	1500
Db	224122	GTTTGTTCCTGCTCGTGGAGACATTCCTTGGAACACCCAGGAATCCACTGGAAGAAAAT	224063
Qy	1501	AGAATTAATACTACAGAGTTTCCACTTGTCTTGAAGAAGAGAGCCCTTCTGAGGGCTG	1560
Db	224062	AGAATTAATACTACAGAGTTTCCACTTGTCTTGAAGAAGAGAGCCCTTCTGAGGGCTG	224003
Qy	1561	TCCATCTAGTGAGATACTATGGAAGAGAGCCTGGAGAGGGCCGAATAAGTGTAGTTGA	1620
Db	224002	TCCATCTAGTGAGATACTATGGAAGAGAGCCTGGAGAGGGCCGAATAAGTGTAGTTGA	223943
Qy	1621	TTACTCATACCTAGAAAGTGCCTTCCCAATTTCTCCAGACCAGCTTGTAGTAAACAAC	1680
Db	223942	TTACTCATACCTAGAAAGTGCCTTCCCAATTTCTCCAGACCAGCTTGTAGTAAACAAC	223883
Qy	1681	GATAGATTATATTTTGGGAGTGCATCCAGTGACCTTGGAAAAACAAGTTCTGATCCAGAAG	1740
Db	223882	GATAGATTATATTTTGGGAGTGCATCCAGTGACCTTGGAAAAACAAGTTCTGATCCAGAAG	223823
Qy	1741	TGAGATTGGGATGAGGAAGCTGAGGATGATGGTTTGTAGTGTAGCTCACTGTCTAGA	1800
Db	223822	TGAGATTGGGATGAGGAAGCTGAGGATGATGGTTTGTAGTGTAGCTCACTGTCTAGA	223763
Qy	1801	CTCAGACCTTGAAACAAGACCCCTGAAGGGCTTCACTTTTGGAACTCTTTCTGCAGTGTAGA	1860
Db	223762	CTCAGACCTTGAAACAAGACCCCTGAAGGGCTTCACTTTTGGAACTCTTTCTGCAGTGTAGA	223703
Qy	1861	TCCTTTAATCCCCAGAACTTTACAGCAACAATTCAGACTGCTGCCAGAAATTTGTTCTCTGA	1920
Db	223702	TCCTTTAATCCCCAGAACTTTACAGCAACAATTCAGACTGCTGCCAGAAATTTGTTCTCTGA	223643
Qy	1921	AGAGCCTTCTGATTCAGAGAAGGATTTGCTGGCAGTCTGATCTAGAGAAATTCCTCCCA	1980
Db	223642	AGAGCCTTCTGATTCAGAGAAGGATTTGCTGGCAGTCTGATCTAGAGAAATTCCTCCCA	223583
Qy	1981	GTCTGGAAGCCCTTCTTGAGACCCTGAGCATAGTTCTGGGAGGGAAGATGACTGGGAATC	2040
Db	223582	GTCTGGAAGCCCTTCTTGAGACCCTGAGCATAGTTCTGGGAGGGAAGATGACTGGGAATC	223523
Qy	2041	TAGTCGAGATGAAGCAGAGAGTCTCAACTGTGGAACCTCAATTCGTAAATTCGTATGAGACC	2100
Db	223522	TAGTCGAGATGAAGCAGAGAGTCTCAACTGTGGAACCTCAATTCGTAAATTCGTATGAGACC	223463

Qy	2101	CTACAACCCCTTTAAATTTTAAAGGCTCTCTTTTCAAAACATCAGGGGGAATAATGAGAAAGGCTG	2160
Db	223462	CTACAACCCCTTTAAATTTTAAAGGCTCTCTTTTCAAAACATCAGGGGGAATAATGAGAAAGGCTG	223440
Qy	2161	TCGTGACTCAAAAGACCCCATCTGAGTCCATTGTGCGCCATTCTCTGAGTGTCAACACTTACT	2220
Db	223402	TCGTGACTCAAAAGACCCCATCTGAGTCCATTGTGCGCCATTCTCTGAGTGTCAACACTTACT	223340
Qy	2221	TTCTTGTGAAGTGCAGCTGTTTGGGGAGCCCAAGAAAGTGAATGTCCAGACTCGGTACACGC	2280
Db	223342	TTCTTGTGAAGTGCAGCTGTTTGGGGAGCCCAAGAAAGTGAATGTCCAGACTCGGTACACGC	223280
Qy	2281	TGAGTTCTTCTCGAGGAGACACACACATGTCARAAAGAAAAGGTAACTTCCT	2337
Db	223282	TGAGTTCTTCTCGAGGAGACACACACATGTCARAAAGAAAAGGTTCCT	223226
RESULT 7			
BC058078		5367 bp	linear
LOCUS			
DEFINITION	Mus musculus protein phosphatase 1, regulatory (inhibitor) subunit 15b, mRNA (cDNA clone MGC:66517 IMAGE:6830215), complete cds.		
ACCESSION	BC058078		
VERSION	BC058078.1	GI:34785997	
KEYWORDS	MGC.		
SOURCE	Mus musculus (house mouse)		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Rodentia; Sciurognathi; Muridae; Mus.		
AUTHORS	1 (bases 1 to 5367)		
	Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G., Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D., Altshul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F., Datchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L., Shapiro,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L., Scheetz,T.E., Brownstein,M.J., Uadin,T.B., Toshiyuki,S., Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J., Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J., McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S., Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W., Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahey,J., Helton,E., Kettman,M., Madan,A., Rodrigues,S., Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y., Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D., Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M., Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smallos,D.E., Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.		
TITLE	Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences		
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)		
MEDLINE	2388297		
PUBMED	12477932		
REFERENCE	2 (bases 1 to 5367)		
AUTHORS	Strausberg,R.		
TITLE	Direct Submission		
JOURNAL	Submitted (08-SEP-2003) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA		
REMARK	NH-MGC Project URL: <a href="http://mgc.nci.nih.gov">http://mgc.nci.nih.gov</a>		
COMMENT	Contact: MGC help desk Email: <a href="mailto:cgaps-remail.nih.gov">cgaps-remail.nih.gov</a> Tissue Procurement: Dr. Jim Lin, University of Iowa cDNA Library Preparation: M. Bento Soares, University of Iowa cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL) DNA Sequencing by: Genome Sequence Centre, BC Cancer Agency, Vancouver, BC, Canada <a href="mailto:info@bcgsc.bc.ca">info@bcgsc.bc.ca</a> Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield, Susanna Chan, Readman Chi, Chris Fjell, Erin Garland, Ran Guin, Letticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven		





1595 GGAGAGGCGGAATAAGTGTAGTCTGATTACTCATACCTAGAGGTGACCTTCCCACTTCT 1654  
 Db |||||  
 1505 GAAGAGGACCACTAATGCACTGACCTCTCAGATAGAGGAGAGGCTTCTGTTTCT 1564  
 Qy |||||  
 1655 GCAGACAGAGCTTGTAGTAACAACTGATAGATTATATTTTGGGAGGTGCATCCAGTGAC 1714  
 Db |||||  
 1565 ACCAGACAG -TTGTAGCAACAACTGATAGATTATATTTTGGGAGGCGCCCACTGAC 1623  
 Qy |||||  
 1715 CTGGAAACAGTCTGTATCCAGAGGTGAGGATTCGGATGAGGAAGCTGAGGATGATGTT 1774  
 Db |||||  
 1624 TTGGAAGCCAGCTCTGATTTCTGAAGTGAAGATTCGGGCGAGGAACCTTGAGGACGATGGC 1693  
 Qy |||||  
 1775 TTTGATAGTATGATGCTCACTGTCAGACTCAGACCTTTGAACAGAGACCTTGAAGGGCTTCAC 1834  
 Db |||||  
 1684 TTTGATAGGATGGCTCCCTGCTGTGATCAGACGTGGAAACAGGACTCGAAGGCCCTTCAC 1743  
 Qy |||||  
 1835 CTTTGGAACTCTTTCTGCAGTGTAGATCCTTATATATCCCAAGAACTTTACAGCAACAAT 1894  
 Db |||||  
 1744 CTTTGGAACTCTTTCCACAGTGTAGATCCTTACAACCCCAAACTTTACAGCCACGATT 1803  
 Qy |||||  
 1895 CAGACTGCTGCCAGAAATTTCTTCCAGAGCCCTCTGATTCAGAGAAAGATTTGCTGGC 1954  
 Db |||||  
 1804 CAGACGGCTGCCAGAAATTTGCCCCAGAGACCCATCAGATTCAGGAGCATCTCTGGTCTGGC 1863  
 Qy |||||  
 1955 AAGTCTGTATCAGAAATTTCTTCCAGTCTGGAAGCCTTCTTGAGACCCCTTGAGCATAGT 2014  
 Db |||||  
 1864 AGCTGTGTAGGGA---GCTGTGAGGAGGACCCCTTCCGAGACCCCGACCATAGT 1920  
 Qy |||||  
 2015 TCTGGGAGGAATGATGATGCGGAATCTAGTCAGATGAAGCAGAGAGTCTCAAACTGTGG 2074  
 Db |||||  
 1921 TCCGGGAGGAATGATGATGCGGAACCGAGTGCAGATGAAGCAGAGAACTCTTAAATTTGTGG 1980  
 Qy |||||  
 2075 AACTCATCTGTAAATCTGATGACCCCTACACCCCTTTAAATTTTAAAGCTCTTTTCAA 2134  
 Db |||||  
 1981 AACTCTTTCTGTCTGATCTGAGACCCCTACACCCCTTTAAATTTTAAAGCTCTTTTCAA 2040  
 Qy |||||  
 2135 ACATCAGGGGAAATGAGAAAGCTGTCTGACTCAAGACCCCATCTGAGTCCACTTGTG 2194  
 Db |||||  
 2041 CGTTCAGGAGAAATTTGNAAGCCGTGAGACTCAAGGCTCTTCTGAGGCTACAGTG 2100  
 Qy |||||  
 2195 GCCATTTCTGAGTGTACACCTTACTTTCTGTAAGGTGAGCTGTTGGGAGGCCAAGAA 2254  
 Db |||||  
 2101 GCCTTCTCTGGCCATCATACCTTACTTTCTTGAAGGCCAGCTGTTAGAGAGCCAAAGAA 2160  
 Qy |||||  
 2255 AGTGAATGTCAGATCCGTACAGCTGACGCTTCTTCTGAGGAGAGACACACATGTC 2314  
 Db |||||  
 2161 GATAATTTGTCAGGCTGTGGGCTGGGTGAGGCTCTTGTGGAAGAAGATACACCCATATC 2220  
 Qy |||||  
 2315 AAAAGAAAAAGGTAACTTCTTGAAGAAAGTTACTGAGTATATATATAGTGGTGTAGAG 2374  
 Db |||||  
 2221 AAGAGAAAAAGGTAACTTCTTGAAGAAAGTTACTGAGTATATATATAGTGGTGTAGAG 2280  
 Qy |||||  
 2375 GATCGAAAGGACCAATGGGAAGAAATTTGCAAGGGATGAGAGGTTCCAGAAACGAAT 2434  
 Db |||||  
 2281 GATCGAAAGGACCAATGGGAAGAAATTTGCAAGGGATGAGAGGTTCCAGAAACGAAT 2340  
 Qy |||||  
 2435 CAAGAAACAGAGATGCTATTGGATTTGCTTGACATTTGACATTTGACACAGAGAAAGATGTT 2494  
 Db |||||  
 2341 CAAGAAACAGAGATGCTATTGGATTTGCTTGACATTTGACATTTGACACAGAGAAAGATGTT 2400  
 Qy |||||  
 2495 AATAGACTCCAGGAAACATGCTTCAAGGACCTTAAATGTTCTCAAGCAATTTGAGTGGC 2554  
 Db |||||  
 2401 AATAGACT---GAGGATCGAGTCAAGGACTTACTGTTGTACAGCAATTTAAGAGTG 2456  
 Qy |||||  
 2555 AGCCTGTAGTCTAGTACATACACTCTTACCTGAGAGGTTCTTTTAAACAA 2614  
 Db |||||  
 2457 A-ACAGCCTGCAACCCGTCGCCACTCTGTCTCTTCTTGAAG- GTTCCCTTTAAACAA 2514  
 Qy |||||  
 2615 ATCTTGGCAGCTGCTCTTTGACATTTTCTTTTATAGAGAAATGTAACCTTGGATCTAGT 2674  
 Db |||||  
 2515 ACATGCGCAGCTGTCTCTTGGACATGTTTCTTAAAGAAACAACTTGTATCTAGAGATGCA 2574

2675 TAATTTTCTTTTTCACATATCCCACTCAGAAACATTCAGGTTTGAAGCCACCTGTA 2734  
 Db |||||  
 2575 TTGATTAATTTTGGTAAATGCTCTCATTAGAA-----ACACCAACTCCGA 2620  
 Qy |||||  
 2735 TAATGAAGAGAACTAGTGTGATTTCTAATCTCCCTCTTTTTCATTTAGTTAGTGTGATGC 2794  
 Db |||||  
 2621 TAATGAAGAAT-----CTCTTATCTGTAATCTCTCTCTTTTCTCTAATTTAGTTAGTGTG 2674  
 Qy |||||  
 2795 TTTT 2798  
 Db |||||  
 2675 GTTT 2678

AC048369 166233 bp DNA linear HTG 25-JUN-2000  
 Homo sapiens chromosome 1 clone RP11-284G5 map 1, WORKING DRAFT  
 SEQUENCE, 29 unordered pieces.  
 AC048369  
 AC048369.2 GI:8705131  
 HTG: HTGS PHASE1; HTGS\_DRAFT.  
 KEYWORDS  
 Homo sapiens (human)  
 ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
 Homo sapiens chromosome 1, clone RP11-284G5  
 Unpublished  
 2 (bases 1 to 166233)  
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,  
 Anderson,S., Baldwin,J., Barna,N., Bastien,V., Bada,F.,  
 Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G.,  
 Campiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,  
 Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,  
 Gage,D., Galagan,S., Gardyna,S., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,  
 Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,  
 Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,  
 Klein,J., Larocque,K., Lamazares,R., Landers,T., Lehoczy,J.,  
 Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,  
 McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,  
 Meldrum,J., Meneus,L., Minova,T., Miranda,C., Mlenga,V., Morrow,J.,  
 Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,  
 O'Neil,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,  
 Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,  
 Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,  
 Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,  
 Teefaye,S., Theodore,J., Tirrell,A., Travers,M., Triggilio,J.,  
 Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,  
 Young,G., Zainoun,J., Zimmer,A. and Zody,M.  
 Direct Submission  
 Submitted (14-APR-2000) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Jun 25, 2000 this sequence version replaced gi:7549699.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html  
 ----- Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIBR  
 Web site: http://www-seq.wi.mit.edu  
 Contact: sequence\_submissions@genome.wi.mit.edu  
 ----- Project Information  
 ----- Project name: L8045  
 Center clone name: 284\_G\_5  
 ----- Summary Statistics  
 Sequencing vector: M13; M7815; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.960731  
 Consensus quality: 151917 bases at least Q40  
 Consensus quality: 158917 bases at least Q30  
 Consensus quality: 161704 bases at least Q20

TITLE  
 JOURNAL  
 COMMENT

Insert size: 178000; agarose-fp  
Insert size: 163433; sum-of-contigs  
Quality coverage: 3.7 in Q20 bases; agarose-fp  
Quality coverage: 4.0 in Q20 bases; sum-of-contigs  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 29 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 1069: contig of 1069 bp in length  
\* 1070: gap of 100 bp  
\* 1170: contig of 1569 bp in length  
\* 2739: gap of 100 bp  
\* 2839: contig of 1305 bp in length  
\* 4143: gap of 100 bp  
\* 4244: contig of 1246 bp in length  
\* 5489: gap of 100 bp  
\* 5590: contig of 1284 bp in length  
\* 6873: gap of 100 bp  
\* 6974: contig of 2180 bp in length  
\* 9154: gap of 100 bp  
\* 9254: contig of 1302 bp in length  
\* 10555: gap of 100 bp  
\* 10656: contig of 3340 bp in length  
\* 13995: gap of 100 bp  
\* 14096: contig of 2594 bp in length  
\* 16690: gap of 100 bp  
\* 16790: contig of 2506 bp in length  
\* 19295: gap of 100 bp  
\* 19396: contig of 3402 bp in length  
\* 22798: gap of 100 bp  
\* 22898: contig of 3556 bp in length  
\* 26454: gap of 100 bp  
\* 26554: contig of 4816 bp in length  
\* 31369: gap of 100 bp  
\* 31370: contig of 4059 bp in length  
\* 35228: gap of 100 bp  
\* 35529: contig of 4903 bp in length  
\* 35629: gap of 100 bp  
\* 40532: contig of 4523 bp in length  
\* 40632: gap of 100 bp  
\* 45154: contig of 5146 bp in length  
\* 45255: gap of 100 bp  
\* 50401: contig of 5857 bp in length  
\* 50501: gap of 100 bp  
\* 56358: contig of 6210 bp in length  
\* 56458: gap of 100 bp  
\* 62667: contig of 6515 bp in length  
\* 62768: gap of 100 bp  
\* 69282: contig of 6888 bp in length  
\* 69383: gap of 100 bp  
\* 76271: contig of 8060 bp in length  
\* 76371: gap of 100 bp  
\* 84431: contig of 7530 bp in length  
\* 84531: gap of 100 bp  
\* 92061: contig of 8951 bp in length  
\* 92161: gap of 100 bp  
\* 10111: contig of 7798 bp in length  
\* 101212: gap of 100 bp  
\* 109009: contig of 11443 bp in length  
\* 109109: gap of 100 bp  
\* 109110: contig of 10047 bp in length  
\* 120552: gap of 100 bp  
\* 120653: contig of 10079 bp in length  
\* 130699: gap of 100 bp  
\* 130799: contig of 15137 bp in length  
\* 130800: gap of 100 bp  
\* 145937: contig of 20197 bp in length.  
\* 146037: gap of 100 bp  
\* 166233: contig of 20197 bp in length.

FEATURES  
source  
1. .166233  
/organism="Homo sapiens"

/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/map="1"  
/clone="RP11-284G5"  
/clone\_lib="RPC1-11 Human Male BAC"  
1. .1069  
/note="assembly\_fragment"  
1170. .2738  
/note="assembly\_fragment"  
2839. .4143  
/note="assembly\_fragment"  
4244. .5489  
/note="assembly\_fragment"  
5590. .6873  
/note="assembly\_fragment"  
6974. .9153  
/note="assembly\_fragment"  
9254. .10555  
/note="assembly\_fragment"  
clone\_end:17  
vector\_side:right"  
10656. .13995  
/note="assembly\_fragment"  
14096. .16689  
/note="assembly\_fragment"  
16790. .19295  
/note="assembly\_fragment"  
19396. .22797  
/note="assembly\_fragment"  
22898. .26453  
/note="assembly\_fragment"  
26554. .31369  
/note="assembly\_fragment"  
31470. .35528  
/note="assembly\_fragment"  
35629. .40531  
/note="assembly\_fragment"  
40632. .45154  
/note="assembly\_fragment"  
45255. .50400  
/note="assembly\_fragment"  
50501. .56357  
/note="assembly\_fragment"  
56458. .62667  
/note="assembly\_fragment"  
62768. .69282  
/note="assembly\_fragment"  
69383. .76270  
/note="assembly\_fragment"  
76371. .84430  
/note="assembly\_fragment"  
84531. .92060  
/note="assembly\_fragment"  
92161. .10111  
/note="assembly\_fragment"  
101212. .109009  
/note="assembly\_fragment"  
109110. .120552  
/note="assembly\_fragment"  
120653. .130699  
/note="assembly\_fragment"  
130800. .145936  
/note="assembly\_fragment"  
146037. .166233  
/note="assembly\_fragment"  
clone\_end:SP6  
vector\_side:right"

ORIGIN

Query Match 44.3%; Score 1303.4; DB 2; Length 166233;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 1304; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY	694	GCTAGATTTTGTCTGGAGTCTACACGCGCCCTGAGAGCCCTGAAGGGACGGGAGAAACGAGC	753
DB	2839	GCTAGATTTTGTCTGGAGTCTACACGCGCCCTGAGAGCCCTGAAGGGACGGGAGAAACGAGC	2898
QY	754	CGCCCCCAGCGGACAGAAATCTTTGAGTTGCTCGACGCTCGATCTCTCAGACCCCTCGGT	813
DB	2899	CGCCCCCAGCGGACAGAAATCTTTGAGTTGCTCGACGCTCGATCTCTCAGACCCCTCGGT	2958
QY	814	CACCAAGTCCCTTCGATTTGGCTAGAGGAGGGGATCCATGGGCAATACCTCGCCCCCAGACCT	873
DB	2959	CACCAAGTCCCTTCGATTTGGCTAGAGGAGGGGATCCATGGGCAATACCTCGCCCCCAGACCT	3018
QY	874	AAAAATTGGAGCTTAAGGCCAAGGGAAGTGCTTTGGACCTCGACACACAGGCTTTTCTCTT	933
DB	3019	AAAAATTGGAGCTTAAGGCCAAGGGAAGTGCTTTGGACCTCGACACACAGGCTTTTCTCTT	3078
QY	934	AGAGCAGCAGCTGTGGGAGTGAGCTGTTGCCCCAGTAGCTTCAATCCCGTCTGTACTC	993
DB	3079	AGAGCAGCAGCTGTGGGAGTGAGCTGTTGCCCCAGTAGCTTCAATCCCGTCTGTACTC	3138
QY	994	TAAACGGGAACTTGGCTCTTCGCCCTCTGGGCCCTCTAAACATTCAAACGCATAGACAAATTT	1053
DB	3139	TAAACGGGAACTTGGCTCTTCGCCCTCTGGGCCCTCTAAACATTCAAACGCATAGACAAATTT	3198
QY	1054	CAGTGTGGTATCCTATTTTGTGAAACCTTTCCTACCTGGACGTCTTTCCTAGGCTAGAAAT	1113
DB	3199	CAGTGTGGTATCCTATTTTGTGAAACCTTTCCTACCTGGACGTCTTTCCTAGGCTAGAAAT	3258
QY	1114	CAGCTATCAGAAACAGTGTGAATACGGAGTAGTCGGCTTCAGACACATTAACCCGAGA	1173
DB	3259	CAGCTATCAGAAACAGTGTGAATACGGAGTAGTCGGCTTCAGACACATTAACCCGAGA	3318
QY	1174	GAGCAGCTGCCCTGAGAGAGGACCAATGTTCATCCCCAGCCGCTGAGTGCAGAACTCATTTCC	1233
DB	3319	GAGCAGCTGCCCTGAGAGAGGACCAATGTTCATCCCCAGCCGCTGAGTGCAGAACTCATTTCC	3378
QY	1234	GGCTCTGTGTGGAGGGATGTCCACCTCTTTCTACGGAAAGGGCTTACAGAAATTCACCAATCT	1293
DB	3379	GGCTCTGTGTGGAGGGATGTCCACCTCTTTCTACGGAAAGGGCTTACAGAAATTCACCAATCT	3438
QY	1294	TCGCATGAACCGGCTGGAAATTCCTTCAACAGGCTTAAACAGGGGCAAGATTTACCCACCCC	1353
DB	3439	TCGCATGAACCGGCTGGAAATTCCTTCAACAGGCTTAAACAGGGGCAAGATTTACCCACCCC	3498
QY	1354	TGACCAAGGATAATGGCTACCA CAGCCTGGAGGGAGAAACACAGCCTTCTCCGGATGGATCC	1413
DB	3499	TGACCAAGGATAATGGCTTACCA CAGCCTGGAGGGAGAAACACAGCCTTCTCCGGATGGATCC	3558
QY	1414	AAAAACATGACAGAGATAACCAACACAGTTGTCTGCTGTGGAGACATTCCTTGAAAA	1473
DB	3559	AAAAACATGACAGAGATAACCAACACAGTTGTCTGCTGTGGAGACATTCCTTGAAAA	3618
QY	1474	CACCCAGGAATCCACTGAAGAAAAAATAGAAATTAATACTACAGAGGTTCCACTTGTCTTT	1533
DB	3619	CACCCAGGAATCCACTGAAGAAAAAATAGAAATTAATACTACAGAGGTTCCACTTGTCTTT	3678
QY	1534	GGAAAGAGAGAGCCCTTCTGAGGGCTGTCCATCTAGTGAGAGATCCTATGAAAAAGGAGCC	1593
DB	3679	GGAAAGAGAGAGCCCTTCTGAGGGCTGTCCATCTAGTGAGAGATCCTATGAAAAAGGAGCC	3738
QY	1594	TGGAGAGGGCCGAATAGTGTAGTTGATTACTCATACCTAGAAAGGTGACCTTCCCATTTTC	1653
DB	3739	TGGAGAGGGCCGAATAGTGTAGTTGATTACTCATACCTAGAAAGGTGACCTTCCCATTTTC	3798
QY	1654	TGCCACAGCCAGCTGTAGTAAACAACTGATAGATTAATTTTGGGAGGTGCATCCAGTGA	1713
DB	3799	TGCCACAGCCAGCTGTAGTAAACAACTGATAGATTAATTTTGGGAGGTGCATCCAGTGA	3858
QY	1714	CCTGGAAACAAGTTCTGATCCAGAAAGGTGAGGATTTGGGATGAGGAAGCTGAGGATGATGG	1773
DB	3859	CCTGGAAACAAGTTCTGATCCAGAAAGGTGAGGATTTGGGATGAGGAAGCTGAGGATGATGG	3918

QY	1774	TTTTGATAGTGTACCTCCTGTCAGACTCAGACCTTGAAACAGACCACTTTACAGGGCTTCA	1831		
Db	3919	TTTTGATAGTGTACCTCCTGTCAGACTCAGACCTTGAAACAGACCACTTTACAGGGCTTCA	3978		
QY	1834	CCTTTGGAACTCTTTCTGCAGTGTAGATCTCTATAATCCCGACGAACCTTTACAGCAACAAT	1893		
Db	3979	CCTTTGGAACTCTTTCTGCAGTGTAGATCTCTATAATCCCGACGAACCTTTACAGCAACAAT	4038		
QY	1894	TCAGACTGCTGCCAAGAATTGTTCTCTGAAGAGCCTTCTGATCTCAGAGAAGGATTGTCTGG	1953		
Db	4039	TCAGACTGCTGCCAAGAATTGTTCTCTGAAGAGCCTTCTGATCTCAGAGAAGGATTGTCTGG	4098		
QY	1954	CAAGCTCTGATCTAGAGAATTCTCCCGAGTCTGGAAGCCTTCCTGA	1998		
Db	4099	CAAGCTCTGATCTAGAGAATTCTCCCGAGTCTGGAAGCCTTCCTGA	4143		
RESULT	9				
AC107842		215781 bp	DNA linear ROD 22-OCT-2003		
LOCUS	Mus musculus chromosome 1, clone RP23-30409, complete sequence.				
DEFINITION	AC107842				
ACCESSION	AC107842.13	GI:37806519			
VERSION					
KEYWORDS	HTG.				
SOURCE	Mus musculus (house mouse)				
ORGANISM	Mus musculus				
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.					
REFERENCE	Birren,B., Nusbaum,C. and Lander,E.				
AUTHORS	1 (bases 1 to 215781)				
TITLE	Mus musculus chromosome 1, clone RP23-30409				
JOURNAL	Unpublished				
REFERENCE	2 (bases 1 to 215781)				
AUTHORS	Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Baxendale,V., Boguslavsky,I., Boukhgalter,B., Brown,A., Camarata,J., Campione,A., Chang,J., Charao,B., Choepel,P., Collangelo,M., Collins,S., Collymore,A., Cook,A., Cooke,P., DeArnellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Gord,S., Govett,M., Graham,L., Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliiev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Larocque,K., Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G., Maclean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrum,J., Meneus,L., Mihova,T., Mieng,V., Murphy,T., Naylor,T., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schuback,R., Seaman,S., Sever,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zemlek,I., Zimmer,A. and Zody,M.				
TITLE	Direct Submission				
JOURNAL	Submitted (24-JAN-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA				
REFERENCE	3 (bases 1 to 215781)				
AUTHORS	Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N., Anderson,M., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T., Boguslavsky,I., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y., Collymore,A., Cook,A., Cooke,P., Corum,B., DeArnellano,K., Diaz,J.S., Dodge,S., Doyle,K., Dorris,L., Erickson,J., Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Iliiev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., Lui,X., Mabbitt,R., Maclean,C., Macdonald,P., Meheus,L., Minova,T., Mieng,V., Murphy,T., Naylor,J., Nelson,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Rayner,C., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schuback,R., Seaman,S., Sever,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zemlek,I., Zimmer,A. and Zody,M.				

Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schuback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Venkataraman, V. S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zlimer, A. and Zody, M.

#### Direct Submission

Submitted (21-OCT-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

#### REFERENCE

##### AUTHORS

Barren, B., Nuebaum, C., Lander, E., Abouelleil, A., Allen, N., Anderson, M., Arachchi, H. M., Barna, N., Bastien, V., Bloom, T., Boguslavsky, I., Boukhgalter, B., Camarata, J., Chang, J., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Corum, B., Dearellano, K., Diaz, J. S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S., Ferrelira, P., FitzGerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagoos, B., Hall, J., Horton, L., Hulme, M., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kellis, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Lui, X., Mabbitt, R., MacLean, C., Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M., Melidrim, J., Meneus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schuback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Venkataraman, V. S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zlimer, A. and Zody, M.

#### TITLE

##### JOURNAL

##### COMMENT

Submitted (22-OCT-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Oct 22, 2003 this sequence version replaced gr:37777388.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WITB  
Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
----- Project Information

Center project name: L20605  
Center clone name: 304\_O\_9

#### FEATURES

##### source

1. .215781  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/chromosome="1"  
/map="1"  
/clone="RP23-30409"  
/clone\_lib="RPCI-23 Female Mouse BAC"

##### misc\_feature

1. .8553  
/note="wgs end extension  
clone end SP6"

##### repeat\_region

complement(2. .143)  
/rpt\_family="URR1A"

##### repeat\_region

441. .478  
/rpt\_family="TTG)n"

##### repeat\_region

776. .815  
/rpt\_family="CA)n"

##### repeat\_region

complement(817. .1006)  
/rpt\_family="ID\_B1"

##### repeat\_region

complement(1092. .1181)  
/rpt\_family="ID3"

##### repeat\_region

1304. .1512  
/rpt\_family="B3"

##### repeat\_region

complement(1754. .1928)  
/rpt\_family="B3"

##### repeat\_region

complement(2578. .2681)  
/rpt\_family="BI\_MM"

repeat\_region  
/rpt\_family="MIR"  
2809. .2907  
/rpt\_family="BI\_MM"  
repeat\_region  
2902. .2957  
/rpt\_family="B2\_Mm1"  
complement(2986. .3154)  
repeat\_region  
/rpt\_family="B3A"  
complement(3608. .3753)  
repeat\_region  
/rpt\_family="BI\_MM"  
complement(3772. .3920)  
repeat\_region  
/rpt\_family="B3A"  
complement(3974. .4059)  
repeat\_region  
/rpt\_family="PB1D10"  
complement(4101. .4207)  
repeat\_region  
/rpt\_family="PB1D9"  
4828. .4971  
/rpt\_family="B1F"  
5328. .5474  
/rpt\_family="BI\_MM"  
complement(6606. .6614)  
unsure  
/notes"<30 qual SINGL region"  
unsure  
complement(6658. .6664)  
/notes"<30 qual SINGL region"  
unsure  
complement(6700. .6708)  
/notes"<30 qual SINGL region"  
unsure  
complement(6737. .6743)  
/notes"<30 qual SINGL region"  
unsure  
complement(6823. .6844)  
/notes"<30 qual SINGL region"  
unsure  
complement(8896. .8900)  
/notes"<30 qual SINGL region"  
7252. .7338  
/rpt\_family="B1F"  
8546. .8671  
/rpt\_family="B1F"  
8554. .8559  
/notes="clone boundary  
clone end:SP6  
site:ECORI"  
9091. .9403  
/rpt\_family="ORR1D"  
repeat\_region  
9406. .9777  
/rpt\_family="ORR1B-int"  
repeat\_region  
9866. .10062  
/rpt\_family="ORR1D"  
repeat\_region  
10732. .10821  
/rpt\_family="ID\_B1"  
complement(11142. .11221)  
repeat\_region  
/rpt\_family="PB1D9"  
12068. .12243  
repeat\_region  
/rpt\_family="B2\_Mm2"  
12341. .12406  
repeat\_region  
/rpt\_family="ORR1D"  
12413. .12778  
repeat\_region  
/rpt\_family="ORR1C"  
complement(13030. .13132)  
repeat\_region  
/rpt\_family="BI\_MM"  
complement(13345. .13408)  
repeat\_region  
/rpt\_family="RSINE1"  
complement(13425. .13690)  
repeat\_region  
/rpt\_family="B4"  
13830. .13980  
repeat\_region  
/rpt\_family="BI\_MM"  
14513. .14624  
repeat\_region  
/rpt\_family="PB1D7"  
14642. .14699  
repeat\_region  
/rpt\_family="G-rich"  
14860. .14963  
repeat\_region  
/rpt\_family="PB1D9"  
complement(15187. .15305)  
repeat\_region  
/rpt\_family="B3"  
complement(15492. .15644)  
repeat\_region  
/rpt\_family="BI\_MM"

Query Match		40.0%;	Score 1175.4;	DB 10;	Length 215781;
Best Local Similarity		72.9%;	Pred. No. 2.7e-306;		
Matches 1642;		Conservative	0;	Mismatches 556;	Indels 53; Gaps 8;
Qy	96	CCATCGCCGAGCCGTGGGGGATCTTCATTCATGGTGTGAAGCGTCGAGCCGACTAGG	155		
Db	151055	CGTCCGCGCGCGGAGGGGTCTTCTCTATGTTGAGCGATCTCACACGGCCTAGG	151114		
Qy	156	GAACCTCTTCCC-CCCCAGGATGAAGTCGATCAGTCGCGCCTATTGCGCGGCTGT	214		
Db	151115	ACGTCTCCTTCCCTAGCCGGGATGACCTAACCGGGTCCGACCGCTTGCGCGGGCCTC	151174		
Qy	215	TCTTCCCTGTGTCTCTGCGCCCGCTGCGCATTTCCGCTGCTGCTGCTGCTTTCTGCTGG	274		
Db	151175	TGGGCGGTCCGCTGACGACTCGTTGCGGAGCCGCGCTCTCTGGGCTCTCTGCGCG	151234		
Qy	275	CTCGAAGATCGCCCTGGAGCAGCGACCGCTGGGCAAGCGCGAGACTCTGTAGGCT	334		
Db	151235	CGCGGAATCGACTCGAGTACCCACTCGTGGCTGGCAAGCGCGAGACTGTGTAGACC	151294		
Qy	335	TCCTCGAATCCGCTGACCTCCAGCGCTGAGCGCGCGGGCCCTACCTGAGAGACTGTC	394		
Db	151295	TCGGATCAGCCCTGCGCTGAGCCGCTGAGCTCTGTCTCTCTCTCTCTGCTGAGAGCCGCC	151354		
Qy	395	AAGAAAAAGGAGATGGAGCCGGGAGCAGCGGATCGCGAAACGGCTTGGCCCTCGGGCG	454		
Db	151355	AAG-GAAGGAGATGGAGCAGGACGACAGCAGGCGCGGAGCGGCTGGCCCTCGGCTG	151413		
Qy	455	GGCTTCCGGTTCGCGCAACCTTTTCCCTGCGGATCGCAAGCAGGCTCTCTTAAGTTC	514		
Db	151414	GGCTCTGGTTCGGGCTGCC---TTCTTTGCGGATCGCAGCGCTGCTCTTGGAGTTC	151470		
Qy	515	CCGACGCTCTTGGCCCGGAAACTCCGGGAACCCCACTGCTTCTCTGCCGAGGCC	574		
Db	151471	CGCGCGCTCTCTCTCGAATAATCCCGGAA-----CTCGGCTCTGCC	151515		
Qy	575	GAGACTCGGGTCAGTTACTGGAAGAACTGCTCTCCAGCTCTTGGCGCGCTCCCGGA	634		
Db	151516	GAGCGTCGGAACCGTACTGGAACAAATGCTTCTGAGCTCTTGGCCCTGCTCCCTAGC	151575		
Qy	635	TTGCTTCAGAGGTCTAAATTTGGAGCCAACTTTTCGTTGGGAATTTTCCGACGATGG	694		
Db	151576	CTATTCCAGAGCTGCTGCTTTGGAGCCAGCTTTTTCGGGGCTGATTTCTTACAGATGG	151635		
Qy	695	CTAGATTTTGTGGAGTCTACAGCGCCCTGAGAGCCCTGAGGACGGAGGAAACCGACC	754		
Db	151636	CTAGATTTTTCGCAAGTTTACAGCGCCCTGAGAGCTTTGAGAGGACGGGGAATCTGCC	151695		
Qy	755	GGCCCCACAGCGCAGAAATCTTTGAGTTTCGCTGAGCTCGACTCTCTCAGACCCCTCGGTC	814		
Db	151696	GCTCCCACGGTGACAGAGTCTCTGAGTTTCACTGCGGCTGGACTCTTCGGAAGACTTGGT	151755		
Qy	815	ACCACTCCCTTGTATGCTAGAGAGGGGATTCACCTGGCAATACCTGCCCGCCAGACCTA	874		
Db	151756	GTCACTTCTTGTATTTGCTAGAGAGGGAATCCAGTGGCACTGCTGCTCTCAGACCTG	151815		
Qy	875	AAATTTGGAGCTTAAGCCCAAGGAAGTCTTTGGACCTCGAGCACAGGCTTTCTCTTA	934		
Db	151816	GAGTTAAACCTAAGGCCCAGGAAGAGCTTTAGACTCTGACGCGCCACATTTCTCTCTG	151875		
Qy	935	GAGCAGCAGCTGTGGGGAGTGAGCTGTTGCCAGTAGCTTTCAATCCCGCTCTGTACTCT	994		
Db	151876	GAGCAGCAGCTGTGGGGAGTGAGTTGTGCCAGTAGCTTTCAAGCTGTGTACTCTCC	151935		
Qy	995	AACCGGGAACCTTGGCTCTTGGCCCTCTGGGCTCTTAAACATTTCAACGATAGACAAATTC	1054		
Db	151936	CACCGAGAACTTGACTCTTCTTCTCTGCGGCTCTGAGCGTTTACAGGCTTAGTAAATTC	151995		
Qy	1055	AGTGTGTATCTATTTGCTGAACCTTCTTACCTGGAGTGTCTTCTCTAGGCTAGAGTTC	1114		
Db	151996	AAGTAGTTTCTCTACTCTGAACTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT	152055		

Qy	1115	AGCTATCAGAACAGTGTGGAATATAGCAGGTAGTCGGCTTTCAGACACTAACCCAGAG	1174		
Db	152056	CGCTGTGAGAGCAGCGCTGGAGGTGGCAGTTTGTGGGTTTCCGAAACACTAACCCAGAG	152115		
Qy	1175	AGCAGCTGCTGAGAGAGGACCATTTGTATCCCGAGCGGCTGAGTGAGAACTCATTTCCG	1234		
Db	152116	AGCTGCTATCTTTCTGAAGATGGTTGTACCCCTCAGCGGTTTGGCGGAGAGATGCGGCA	152175		
Qy	1235	GCCTCGTGGCAGGATGTCCACCTTTTCTACGGAAGGCTTACCAGAAAATTCACCATCTT	1294		
Db	152176	ACCGCTTGGAGAGGTGTCCGCTCTCTCTACAGAGCGCTTCCGGAATCCACACCTT	152235		
Qy	1295	CGCATGAAAACGGCTGGAAATTCCTTCAACAGGCTTAAAGGGGCAAGATTTTACCCACCCCT	1354		
Db	152236	CGTATGAAAACGGCTAGAAATTCCT---CCAGGCTTAAAGGCAAGAGTTTACCCACCCCT	152292		
Qy	1355	GACAGGATATGCTTACCAAGCTGAGGAGGAAACACAGCCTTCTCCGATCGGATGCCA	1414		
Db	152293	GACCAAGATATGCTTATCATAGCTTGGAGGAAACATAAACCCTTCTCCGATGGACCCA	152352		
Qy	1415	AAACACTGACAGATACCCCAACACAGTTTGTCTCTGCTGAGACATTTCTCTGGAAC	1474		
Db	152353	CAACRTTGACAGATTAACCCAGACAGCGGTGTCCCTGCTGACAGACAGGCC-----	152405		
Qy	1475	ACCAGGAATCCACTGAAGAAAAAATAGAAATTAATACTACAGAGTTTCCAATTGCTTTG	1534		
Db	152406	-----GGAGCCCACTGAGAAAAAACAGAAATTTGGTGATTCAGAAAGTTTCAC-----	152452		
Qy	1535	GAAGAACAGAGCCCTTCTGAGGCTGTCCATCTAGTGAATACCTATGGAAGAGGAGCCT	1594		
Db	152453	-----AGAGCCCGCCAGGGAAGCAGTCTGTTTTGTGAATTAACCCGTGGAAAAAGATGT	152505		
Qy	1595	GGAGAGGCGCGAATAAGCTGTAGTTGATTACTCATACCTAGAAGGTGACTTCCCATTTCT	1654		
Db	152506	GAAGAGACCACTAATGCACTGACCTCTCAGATAGAGAGAGAGGCTTCTGTTTCT	152565		
Qy	1655	GCCAGACAGGCTTGTAGTAAACAACTGATAGATTATATTTTGGAGGTGCAATCAGTGAC	1714		
Db	152566	ACCAGACCGATTCTGACCAACAACTGATAGATTATATTTTGGAGGCGGCCCACTGAC	152625		
Qy	1715	CTGGAACAAGTTCTGATCAGAAAGGTGAGGATTTGGATGAGGAAGCTGAGGATGATGT	1774		
Db	152626	TTGGAACCGACTCTGATCTGAAAGTGAGGATTTGGGGCGAGGAACTCTGAGGACGATGGC	152685		
Qy	1775	TTTCATAGTATAGTCACTGCTCAGACTCAGACCTTGAAACAAAGACCTCTGAAGGCTTTCAC	1834		
Db	152686	TTTGTATGAGATGCTCTCTCTCTGTAATCAGACGTGGAAACAGGACTCGGAAGGCTTTCAC	152745		
Qy	1835	CTTTGGAACTCTTTCTGAGGTAGATCTTATTAATCCCAGAACTTTTACAGCAACAATTT	1894		
Db	152746	CTTTGGAACTCTTTTCCACAGTGTAGATCTTACAAACCCCAAAACTTTTACAGCCACGATTT	152805		
Qy	1895	CAGACTGCTGCCAGAAATTTTCTGAAGAGCTTCTGATTTCCAGAGAGGATTTTCTCTGGC	1954		
Db	152806	CAGACGCTGCCAGAAATTTGCCCGCCAGAGACCCATCAGATTACAGGGAATCTCTGCTGGC	152865		
Qy	1955	AAGTCTGATCTAGAGAAATTTCTCCAGTCTGGAAGCCTTCTCTGAGACCCCTGAGCATAGT	2014		
Db	152866	AGCTGTGTTAGGGA---GCTGTGAGAGGAGCCCTTCCGAGAGACCCCGACCATAGT	152922		
Qy	2015	TCTGGGAGGAGATGATGATGGGAAATCTAGTGCAGATGAAGCAGAGAGTCTCAAACTGTGG	2074		
Db	152923	TCCGGGAGGAGATGATGATGGGAAACCGAGTGCAGATGAAGCAGAGAAATCTTAAATTTGG	152982		
Qy	2075	AACTCATTTCTGTAATCTGATGACCCCTTCAACCTTTTAAATTTTAAAGGCTCTCTTTCAA	2134		
Db	152983	AACTCTTCTGTCTAATCTGAGGACCCCTTCAACCTTTTAAATTTTAAAGGCTCTCTTTCAA	153042		
Qy	2135	ACATCAGGGGAAAAATGAGAAAGGCTGTGCTGACTCTCAAGAGCCCATCTGTGATCCATTTGT	2194		
Db	153043	CCGTGAGGAGAAATTTGGAAGGCGCTCAGGACTCAAGAGGCTCTCTCTGAGGCTACAGTG	153102		
Qy	2195	GCCATTTCTGAGTGTACACCTTACTTTCTTTGAGGTGCAGCTGTTGGGAGCCACAGA	2254		



```

||||| 153103 GCCTTCTCTGCGCATCATACCTTACTTCTTGTGAAGCCCGAGCTGTGTAGAGGCAAGAA 153162
||||| 2255 AGTCAATGTCTCAGACTCGGTACAGCGTGACGCTTCTTCTTGAGGAGCAACACACATGTC 2314
||||| 153163 GATTAATTGTCTCAGGCTGTGGCTGGTGAGGCTTCTTGCGAGAGATACACCCATATC 153222
||||| 2315 AAAAGAAAAAGGTAACCTTCTTGAAGAAG 2345
||||| 153223 AAGAGAAAAAGGTTTGTCTTCTTTACATG 153253

RESULT 10
AC128396/c
LOCUS
DEFINITION Rattus norvegicus clone CH230-194B23, *** SEQUENCING IN PROGRESS
***. 6 unordered pieces.
ACCESSION AC128396
VERSION AC128396.4 GI:23322373
KEYWORDS HTG; HTGS PHASE1; HTGS DRAFT; HTGS_ENRICHED.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 200490)
Muzny D,Marie., Metzker M, Lee., Abramzon S., Adams C., Alder J.,
Allen C., Allen H., Alabrooks S., Amin A., Anguiano D.,
Anyalebechi V., Aoyagi A., Ayodeji M., Baca E., Baden H.,
Baldwin D., Bandaranaike D., Barber M., Barnstead M., Benahmed F.,
Blawie K., Blair J., Blankenburg K., Blyth P., Brown M.,
Bryant N., Buhay C., Burch P., Burrell K., Calderon E.,
Cardenas V., Carter K., Cavazos I., Cesar H., Center A.,
Chacko J., Chavez D., Chen G., Chen R., Chen Y., Chen Z., Chu J.,
Cleveland C., Cockrell R., Cox C., Coyle M., Cree A., D'Souza L.,
Davila M.B., Davis C., Davy-Carroll L., De Anda C., Dederich D.,
Delgado O., Denson S., Deramo C., Ding Y., Dinh H., Divya K.,
Drapac H., Dugan-Rocha S., Dunn A., Durbin K., Duval B., Eaves K.,
Egan A., Escotto M., Eugene C., Evans C.A., Falls T., Fan G.,
Fernandez S., Finley M., Flagg N., Forbes L., Foster M., Foster P.,
Fraser C.M., Gabisi A., Ganta R., Garcia A., Garner T., Garza M.,
Gebregorgie E., Geier K., Gill R., Grady M., Guerra W., Guevara W.,
Gunaratne P., Haaland W., Hamill C., Hamilton C., Hamilton K.,
Harvey Y., Havlak P., Hawes A., Henderson N., Hernandez J.,
Hernandez B., Hines S., Hladun S.L., Hodgson A., Hogues M.,
Hollman R., Howells S., Hulyk S., Hume J., Idlebird D., Jackson A.,
Jackson L., Jacob L., Jiang H., Johnson B., Johnson R., Jolivet A.,
Karpathy S., Kelly S., Kelly S., Khan Z., King L., Kovar C.,
Kowis C., Kraft C.L., Lebow H., Levan J., Lewis L., Li Z., Liu J.,
Liu J., Liu W., Liu Y., London P., Longacre S., Lopez J.,
Lorenshewa L., Loulseghe H., Lozado R.J., Lu X., Ma J.,
Maheshwari M., Mahindartne M., Mahmoud M., Malloy K., Mangum A.,
Mangum B., Mapus P., Martin K., Martin R., Martinez E.,
Mawhinney S., McLeod M.P., McNeill T.Z., Meenen E.,
Milosavljevic A., Miner G., Minja E., Montemayor J., Moore S.,
Morgan M., Morris K., Morris S., Munidasa M., Murphy M., Nair L.,
Nankervis C., Neal D., Newton N., Nguyen N., Norris S.,
Nwaekeleneh O., Okwono G., Olarnpunsagoon A., Pal S., Parks K.,
Pasternak S., Paul H., Perez A., Perez J., Pfannkuch C.,
Plopper F., Poindester A., Popovic D., Primus E., Pu L.-L.,
Puzo M., Quiroz J., Rachlin E., Reeves K., Regier M.A., Reigh R.,
Reilly B., Reilly M., Ren Y., Reuter M., Richards S., Riggs F.,
Rives C., Rodkey T., Rojas A., Rose M., Rose R., Ruiz S.J.,
Sanders W., Savary G., Scherer S., Scott G., Shatsman S., Shen H.,
Shetty J., Shivatsbeyn A., Slason I., Sitter C.D., Smajda D.,
Sneed A., Sodergren E., Song X.-Z., Sorelle R., Sosa J.,
Steimle M., Strong R., Sutton A., Svatek A., Taber P., Taylor C.,
Taylor T., Thomas N., Thomas S., Tingey A., Trejos Z., Usmani K.,
Valas R., Vera V., Villanasa D., Waldron L., Walker B., Wang J.,
Wang Q., Wang S., Warren J., Warren R., Wei X., White F.,
Williams G., Willson R., Wleczyk R., Woodman H., Worley K.,
Wright D., Wright R., Wu J., Yakub S., Yen J., Yoon L., Yoon V.,
Yu F., Zhang J., Zhou J., Zhou X., Zhao S., Dunn D., von

```

```

Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,
Weinstock, G. and Gibbs, R.A.
Unpublished
2 (bases 1 to 200490)
Worley, K.C.
Direct Submission
Submitted (19-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 200490)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (26-SEP-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Sep 26, 2002 this sequence version replaced gi:23771040.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). As a result, the
sequence may extend beyond the ends of the clone and there may be
contigs that consist entirely of whole genome shotgun sequence
reads. Both end sequences and whole genome shotgun sequence only
contigs will be indicated in the feature table.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GYZL
Center clone name: CH230-194B23
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 131649 bases at least Q40
Consensus quality: 136220 bases at least Q30
Consensus quality: 138894 bases at least Q20
Estimated insert size: 161200; sum-of-contigs estimation
Quality coverage: 4x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 21048: contig of 21048 bp in length
* 21049 21148: gap of unknown length
* 108746 108746: contig of 87598 bp in length
* 108747 108846: gap of unknown length
* 108847 161819: contig of 52973 bp in length
* 161820 161919: gap of unknown length
* 161920 176640: contig of 14721 bp in length
* 176641 176740: gap of unknown length
* 176741 187347: contig of 10607 bp in length
* 187348 187448: gap of unknown length
* 187448 200490: contig of 13043 bp in length.
FEATURES
Location/Qualifiers
1..200490
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-194B23"
1..1226
/notes="wgs contig"
13967..15951
/notes="wgs contig"
15951..17948
/notes="wgs contig"
17948..21048
/notes="wgs contig"
misc_feature
misc_feature
misc_feature

```







```
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 35084: contig of 35084 bp in length
* 35085 35184: gap of unknown length
* 35185 224939: contig of 189755 bp in length
* 224940 225039: gap of unknown length
* 225040 227054: contig of 2015 bp in length.
*
FEATURES             Location/Qualifiers
     source            1..227054
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-191M14"
     misc_feature      145242..147317
                     /note="wgs_contig"

ORIGIN
Query Match      38.7%; Score 1137.4; DB 2; Length 227054;
Best Local Similarity 71.6%; Pred. No. 5.4e-296;
Matches 1711; Conservative 0; Mismatches 586; Indels 92; Gaps 13;

QY 1 ATTCTGGGCTTCGCTTCACCGCAGCCAGCGGGCTTACCGAGTCTCTTCGGGTATCCGGTTG 60
DB 153007 ATTTTGAAGTTCCGTTCTCGCTCTCGCCGCCACACAGCCCTTAACGCTCTGTGTATT 152948

QY 61 CTCAGGGGCTTTTCAACCTCTGTCTAGTCGGAACACCATCGCGAGGCGCTGGGGGACT 120
DB 152947 GCG-GCCTCCGTTGGGCGCGGTCTTCGGAACACGTTGCGCGCGCGCGCTGTGAGGGT 152889

QY 121 CTTATTCATGTGTTGAAGCGTCGAGCGGACTTAGGGAACCTCTTCCCGCGCAGGATGGA 180
DB 152808 CTTCTCTATGTGGAGCGGTCTCACAGGACTAGGCGGCTCCATTCTCGCGGAGATGA 152829

QY 181 AGTCGATCAGTGGCGGCTATTGGCGGGGCTGTTCTTCCCTGTGTTCTGGCGCGCGTG 240
DB 152828 CGTAGTCGGGTGCGCACCGCTTGGCGAGCGCGCGCGGTACGGTATCATCGCTCGCTA 152769

QY 241 CGCATTCGTCGCCCTCTGTGGCTTTCTGTGGCTCGAAGATCGGCTGAGAGAGCCAC 300
DB 152768 CGGAAGCGCGCGCTCTGTGGCC-----CTGTGGCGCGGGAATCGGAATCAGTACCCAC 152712

QY 301 GCACCGCTGGGCAAGCGCGAGACTCTGTAGGCTTCTCCGAATCCCGTCCGACCTCCAGC 360
DB 152711 TCCGTGTCTGGCAAGGCTTAAGACTGTGTAGG---CCTCGGATCCTGCGCGCTCTGACGC 152655

QY 361 CGCTAGCGCGCGGCGCTTACTGAGAGACTCTGAAGAAAGAGATGGAGCGGGGAC 420
DB 152654 CGCTAGCTCCGTACTCTGTGCGCGGAGAGTTGCCAAGG-AAAGGAGATGGAGACGGGAAC 152596

QY 421 AGGCGGATCGGGAACGCTTGGCGCTCGGCGGGCTTCCGGTCTTGCGCCACCCCTTTT 480
DB 152595 ACACAGGGCTCGGAAGCGGCTTACTCTCGGATGGGCTCCTGGTTCGGGCTGCGCC---TT 152539

QY 481 CCTTCGGGATCGCAAGCAGGCTCTTCTAAGTTCGCCGACGCTCTTGGCCCGGAAAACTC 540
DB 152538 CTTTCGGGTTCCGACGCGCTGACCTTCGAGTTCCCGAAGCCTTCTTCTCGACCAAACTCT 152479

QY 541 CGGGAACCCCACTGCTTCTCTGCCAGCGCGGAGCTGGGTCAGTCTACTGACGAA 600
DB 152478 CGAG-----GTCCTCGCGCGCGCGGAGCTCGGACCGGTACTGGAACCA 152434

QY 601 ACTGCTCCCAAGCTCTTTCGCGCGCTCCCGGATTTGCTTCAGAGGTGCTTAATTGGAG 660
DB 152433 ATTGTTTCCCAAGCTCTTTCGCGCTCTCCCTAGCTTGTTCAGAGAGTCTTAATTGGAG 152374

QY 661 CCAACTTTTCGGTGAATGTTTCCGACGAGATGGCTAGATTGTTGCTGGAGTCTACAGCGC 720
```

```
DB 152373 TCAGCTTTTCGGTGGCATGATTCGACAGATGTTAGATTTTTCAGCAAGTTACAGCGC 152314
QY 721 CCTGAGAGCCTGAAGGAGCGGGAACACCGCGCCACACAGCGCAGAAATCTTTGAG 780
DB 152313 TCTAAGAGCTTTGAGAGGAGCGGAGGAATCAGCGCTCCACACAGTGCAGAAATCTTTGAA 152254
QY 781 TTGCTCTCAGCTCGAC---TCCTCAGACCCCTCGGTCAACAGTCCCTTGAATTCGCTAGA 837
DB 152253 TTCTCTGGGCTGGACTCTTCTTGGGAAGACTTGGTGTCACTTCTTGTATTCGCTAGA 152194
QY 838 GGAGGGATCCATGGGCAATACTGCGCCCGCAGACTAAATTTGAGGTTAAGGCCAAGG 897
DB 152193 GGAGGGACTCCAGGGGCACTGCTCGTCTCAGACTCGAAATGGAATTCAGAGGCCAGGA 152134
QY 898 AAGTGTCTTGGACCTCGCAGCAGAGGCTTTCTCTTAGCAGCAGCTGTGGGAGTGA 957
DB 152133 AAGAGCTTTTAGACTCTACAGCGCACACTTTCTCTTGGAGCAGCAGCTGTGGGAGTGA 152074
QY 958 GCTGTGCCAGTAGCTTCAATCCCGTCTGTACTCTAAACCGGAACTTGGCTCTTCGCC 1017
DB 152073 GTTGTGCGCGGTAGTCTTCAACCGGTCTAGTCTCCACCGAGAACTTGGCTCTCATC 152014
QY 1018 CTCTGGGCTCTAAACATTCACGCAATTCAGCTAGTGTGTATCTTATTTGCTGAA 1077
DB 152013 CTCTGGGCTCTGAGTACTCAGAGCTTGGGCAATTTCAAGGTAGTCTTCTATCTCTGAA 151954
QY 1078 CCCTTCTTACTCGGACTGCTTCTTAGGCTAGAGTCACTATCAGAACAGTGTGTAAGAA 1137
DB 151953 CCCTTCTTACTCGGACTGCTTCTTCCAGTGTAGAGTCCGCTGTGAGAACAGCTGCGAGG 151894
QY 1138 TAGCAGGTAGTGTGGCTTCCAGACATTAACCCAGAGAGCAGCTGCTGAGAGAGGACCA 1197
DB 151893 TGGCCAACTTGTGATTTCCGAACACTATCCCCAGAGAGCTGCTACCTTCTGAAGATGC 151834
QY 1198 TTGTCACTCCAGCGCTGAGTGAGAACTCATTTCCGGCTCTGTGGCAGGATGTCACC 1257
DB 151833 TTGTCACTCCAGCGCTGTGTGTGAGATGTGCGCAACTGCTGTGGCGAGGTGTCGCC 151774
QY 1258 TCTTTCTACGGAAGGCTTACCAGAAATTCACCACTCTTTCGATGAAACGGCTGGAATTCCT 1317
DB 151773 CCTCTACAGAGGCTTCCAGAAATTCACCACTCTTTCGATGAAACAGCTAGAAATTCCT 151714
QY 1318 TCAACAGGCTAAACAAGGGGCAAGATTTACCCACCCCTGACAGGATTAATGCTACCAAG 1377
DB 151713 T---CAGGCAAAACAAGGGGCAAGATTTACCCACCCCTGACCAAGCAATGCTATCATAG 151657
QY 1378 CTGGAAGAGAACACAGCCTTCTCGGATGGATTCAAAACACCTGCGAGAGATAACCCAC 1437
DB 151656 CTGGAAGAGAGACACTGCTTCTCGGATGGATTCGCAACATTTGCAAGATAAGGCGC 151597
QY 1438 ACAGTTTCTCTGCTCTGAGACATTCCTGGAAACACCCAGGAATCCACTGAAGAA 1497
DB 151596 ACAGGAGTCTCCCTCTGAG-----CAGGCCGCGCGCTGAGAA 151549
QY 1498 AATAAGATTTAATTAATAAGAGGTTCCACTTGTGTTTGGAAAGAGAGAGCCTTCTGAGG 1557
DB 151548 AGTAGAATGTTGTTGAGTTCAGGAAGTTTAC-----AGAGCCCGGAGGAG 151504
QY 1558 CTGTCCATCTAGTGAGATACCTATGGAAGAGGAGCTTGGAGGCGGATAGTGTAGT 1617
DB 151503 CAGTCTCTTGTGAGTTAAGTGTGCAAAAGAAATGTGAA----- 151464
QY 1618 TGATTACTCATACCTAGAGAGTGTACCTTCCATTTCTGCGCAGACAGCTTGTAGTAACA 1677
DB 151463 -GACTACTCAGACATAGGGGTAACTTCCAGTCTCTACAGACAGCAGCTGTGACCAACA 151405
QY 1678 ACTGATAGATTATATTTTGGAGGTGCATCCAGTGAACCTGGAACAAGTCTGTATCAGA 1737
DB 151404 ACTGATAGATTATATTTTGGAGGAGCCTCAGTGAATTTGGAACACAGCTCCGATCTGA 151345
QY 1738 AGGTGAGATTGGGATGAGAGCTGAGGATGATGGTTTGTAGTATGATGATGATGATGATG 1797
```

Db	151344	AAGTGAGGATGGGATGAGGAAACCTGAGGATGATGTTTGGATAGTGATGCTCCCTATC	151285
Qy	1798	AGACTCAGACCTTGAACAAGACCTGAGGCTTCACTTTGGAACTCTTTCTGCAGTGT	1857
Db	151284	TGAATCAGACAGGAGACAGACTCGAAGGCTTCACTTTGGAACTCTTTCTACAGTGT	151225
Qy	1858	AGATCTTATATATCCCGAAGACTTTTACAGCAACAAATTCAGACTCTGCCGAATTTGTTCC	1917
Db	151224	AGATCTTATACAAACCCCAAGACTTTTACAGCCACAATTCAGACTCTGCCGAATTTGTTCC	151165
Qy	1918	TGAAGAGCCTTCTGATTCAGAGAGGATTTGTCTGGCAAGTCTGATTCAGAGATTTCTC	1977
Db	151164	CAGAGACCAATCTGATTCAGAGAGGATTTGTCTGGCAAGTCTGATTCAGAGATTTCTC	151108
Qy	1978	CCAGTCTGGAAGCTTCTCTGAGACCCCTGAGCAGTGTCTGGGAGGAGGATGACTGGGA	2037
Db	151107	TCAGGCCACACTCTTCTCCAGACCCCTGAGCAGTGTCTGGGAGGAGGATGACTGGGA	151048
Qy	2038	ATCTAGTGCAGATGAAGCAGAGAGTCTCAAACTGTGGAATCATTTCTGTATTTCTGATGA	2097
Db	151047	ATCCAGTGCAGATGAAGCAGAGAGTCTCAAACTGTGGAATCATTTCTGTATTTCTGATGA	150988
Qy	2098	CCCTACACCCCTTAAATTTTAAAGTCTCTTTTCAACATCAGCGGAAATGGAAGG	2157
Db	150987	CCCTACACCCCTTAAATTTTAAAGTCTCTTTTCAACATCAGCGGAAATGGAAGG	150928
Qy	2158	CTGTGCTGACTCAAGACCCCATCTGAGTCCATTTGTGCCATTTCTGAGTGTCAACCTT	2217
Db	150927	CAGTC-----AGGATCTCTGAGGCCACAGTGTCTTTCTGGCATCAGACTT	150877
Qy	2218	ACTTTCTTGTAAAGTGCAGCTGTGGGAGCCAAAGTGAATGTCCAGACTCGGTACA	2277
Db	150876	ACTTTCTTGTGAGGCACAACCTGTTAGAGAGCCAAAGTGAATGTCCAGACTCGGTACA	150817
Qy	2278	GCGTGAGCTTCTTCTGAGGAGACACACACATGTCAAGAAAGGTAACCTTCT	2337
Db	150816	GGGTGAGGCACCTTCTGAGGAGACACACACATGTCAAGAAAGGTAACCTTCT	150757
Qy	2338	TGAAGAAGTACTCAGTATATATAAGTGGTGGATGAGATCGCAAGGA	2386
Db	150756	TTTACATGACTTAAATTTGTTGTTATTTGTTATGTTACGTAGTAGAGA	150708

RESULT 12  
AC105827  
LOCUS  
DEFINITION Rattus norvegicus clone CH230-101J7, \*\*\* SEQUENCING IN PROGRESS

AC105827  
VERSION HTG; HTGS PHASE2; HTGS DRAFT; HTGS\_ENRICHED.  
KEYWORDS Rattus norvegicus (Norway rat)  
SOURCE Rattus norvegicus  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE  
AUTHORS  
1 (bases 1 to 248458)  
Muzny, D. Marie, Metzker, M. Lee, Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Alibrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Cesari, H., Chen, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, I., Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denison, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,

Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorenshehwa, L., Loulseghe, H., Lozano, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mathew, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, C., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwaokemele, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Polndexter, A., Popovic, D., Primus, E., Pu, L., L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajd, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Soosa, J., Steinle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villaseana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wlezyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

Direct Submission  
Unpublished  
2 (bases 1 to 248458)  
Worley, K.C.  
Direct Submission  
Submitted (10-JAN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 248458)  
Rat Genome Sequencing Consortium.  
Direct Submission  
Submitted (21-SEP-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
On Sep 21, 2002 this sequence version replaced gi:21736987.  
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). As a result, the sequence may extend beyond the ends of the clone and there may be contigs that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)  
----- Project Information  
Center project name: GNVL  
Center clone name: CH230-101J7  
----- Summary Statistics  
Assembly program: Phrap; version 0.990329  
Consensus quality: 231489 bases at least Q40  
Consensus quality: 232351 bases at least Q30  
Consensus quality: 232940 bases at least Q20  
Estimated insert size: 249328; sum-of-contigs estimation  
Quality coverage: 5x in Q20 bases; sum-of-contigs estimation  
-----  
\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft.html)).



```
Db 5595 ACTGATAGATTATATTTTGGGAGGAGCCTCCAGTCACTTGGACACCAGCTCCGATTCCTGA 5654
Qy 1738 AGGTGAGGATTTGGGATGAGGAAGCTGAGCATGATGTTTTCATAGTATGATCACTGTC 1797
Db 5655 AAGTGAGGATTTGGGATGAGGAAGCTGAGCATGATGTTTTCATAGTATGATCACTGTC 5714
Qy 1798 AGACTCAGACCTTGAACAAGACCCCTGGAAGGCTTCACCTTTTGGAACTCTTTCTCGAGTGT 1857
Db 5715 TGAATCAGACAGGGAACAAGACTCCGAAGGCTTCACCTTTTGGAACTCTTTCTCGAGTGT 5774
Qy 1858 AGATCCTTATATATCCCAAGACTTTACGCAACATTCAGACTCTGCGCAATTTGCTCC 1917
Db 5775 AGATCCTTATATATCCCAAGACTTTACAGCCACATTCAGACTCTGCGCAATTTGCTCC 5834
Qy 1918 TGAAGAGCCTTCTGATTCAGAGAAGGATTTCTGCGCAAGTCTGATCTAGAGAATTCCTC 1977
Db 5835 CAGAGACCCATCTGATTCAGAGAAGTCTGCTGCTGCGCAACTCTGATGTAGGAGT---TC 5891
Qy 1978 CCAGTCTGGAAGCCTTCTGAGACCCCTGAGCATGATTTCTGGGAGGAAGATGCTGGGA 2037
Db 5892 TCAGGCCACACTCTCTCCCAAGACCCCTGACAGAGTTCCGGGAGGAGGATGACTGGGA 5951
Qy 2038 ATCTAGTGCATGACGAGCAGAGTCTCAAACTCTGGAACCTCATCTCTAATTCCTGATGA 2097
Db 5952 ATCCAGTGCATGAGCAGAGAATCTTAGATTGTGGAACCTCTTTCTGTAAATCTGAGGA 6011
Qy 2098 CCCCTACAACCTTTAAATTTTAAAGGCTCTCTTTTCAAACTCAGGGGAAAATGAGAAAGG 2157
Db 6012 CCCCTACAACCTTTAAATTTTAAAGGCTCTCTTTTCAAACTCAGGGGAAAATGAGAAAGG 6071
Qy 2158 CTGTCTGAGTCAAGACCCCATCTGAGTCCATTTGTGSCCATTTTCTGAGTGTACACCTT 2217
Db 6072 CAGTC-----AGGCATCTCTGAGGCCACAGTGGTCTTTCTGGGCATCAGACCTT 6122
Qy 2218 ACTTCTGTGAAGTGCAGCTGTGGGAGGACCAAGATGAATGTCAGACTCGGTACA 2277
Db 6123 ACTTCTGTGAGGACCAACTGTGTAGAGGACCAAGATGAATGTCAGACTCGGTGCT 6182
Qy 2278 CGGTGACCTTCTTCTGAGGAAGACACACACATGTCAAAAGAAAAGGTAACCTTCCT 2337
Db 6183 GGGTGAAGCATTTCTGAGGAAGATACACCATATCAGAGAAAAGGTTTGTTCCTCA 6242
Qy 2338 TGAAGAAGTTACTGAGTATATATAAGTGGTGTGAGGATCGCAAGA 2386
Db 6243 TTTACATGACTTATAATTTGTTGTTATTTGATATGTTTACAGTAGAAGA 6291

RESULT 13
AC119774
LOCUS Rattus norvegicus clone CH230-146P11, WORKING DRAFT SEQUENCE, 2
DEFINITION AC119774 212261 bp DNA linear HTG 19-NOV-2002
unorderd pieces.
ACCESSION AC119774
VERSION AC119774.6 GI:25095262
KEYWORDS HTG; HTGS PHASE1; HTGS DRAFT; HTGS_FULLTOP.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 212261)
Muzny,D,Marie., Metzker,M, Lee., Abramson,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alabrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Bigswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burk,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
```

Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falle,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Garbogeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jarkson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpthy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuewa,L., Loulseghe,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaokelimeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Polindexter,A., Popovic,D., Primus,E., Pu,L., L., Puzo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartabeyn,A., Sisson,I., Sitter,C.D., Smaja,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steinle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,R., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villaeana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausen,A., Weiles,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.

Direct Submission  
Unpublished  
2 (bases 1 to 212261)  
Worley,K.C.  
Direct Submission  
Submitted (02-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 212261)  
Rat Genome Sequencing Consortium.  
Submitted (19-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
On Nov 19, 2002 this sequence version replaced gi:23194778.  
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)  
----- Project Information  
Center project name: GVQB  
Center clone name: CH230-146P11  
----- Summary Statistics



```

Assembly program: Phrap; version 0.990329
Consensus quality: 198836 bases at least Q40
Consensus quality: 201017 bases at least Q30
Consensus quality: 202375 bases at least Q20
Estimated insert size: 203079; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.
*
* 1 11068: contig of 11068 bp in length
* 11069 11168: gap of unknown length
* 11169 212261: contig of 201093 bp in length.
FEATURES             Location/Qualifiers
     source            1..212261
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-146P11"
     misc_feature      1..1217
                     /note="wgs_contig"
ORIGIN
Query Match      38.1%; Score 1120.2; DB 2; Length 212261;
Best Local Similarity 71.6%; Pred. No. 2.4e-291;
Matches 1771; Conservative 0; Mismatches 603; Indels 99; Gaps 19;

QY 347  CGTGACCTCCAGCGCTGAGCGCGCGGCCCTACTGAGACATGTGCAAGAAAAGGAG 406
Db 140598  CTTGCGTCTGATGCGCGCTGAGCTCGGTACTCTGCGCGGTGAAGTTGCCAAG- 140656

QY 407  ATGGAGCGGGACAGCGGATCGCGGAAACGGCTTGGCCCTCGCGCGGCTTCGGTTTC 466
Db 140657  ATGGAGACGGGAACACAGGGGCTC-AMAGCGCGCTACTCTCGGATGGGCTCTGATTA 140715

QY 467  TGGCCACACCTTTTCCCTCGGCGATGCGAAGAGGCTCTTTAAAGTTCCCGACGCTCTT 526
Db 140716  TGCTGCGCTTCTCTAG---GGCGATCACAGCGCTGACCTTCTGAGTTCCGAAACCTTCT 140772

QY 527  GGCCCGGAAAACTCGCGGAAACCCACATGCTTTCTGCGCCAGCGAGACTCGGTC 586
Db 140773  TCTCAACCAATCTCGAGG-----CTCTGCTGGTCCGAGCCTCAGACC 140817

QY 587  AGTTACTGGACGAACATGCTCTCCAGCTCCTTGGCGCGCTCCCGGATTTGCTTCAGAG 646
Db 140818  AGGTACTGGACCAATATGCAATCCAGATTCCTTG---CTCTCCCTAGCTTGTCTCAGAG 140874

QY 647  GTGCTAAATTTGGAGCAACTTTTTCGGTGGAAATGTTTCCGACAGATGGCTAGATTTGCT 706
Db 140875  CTGCTTAATTTGGAGTCAGCTTTTCGGTGGCATGATCTGACCATGGCTATACCTTTTCA 140934

QY 707  GGAGTCTACAGCGCCTGAGAGCCCTGAAGGGACGGGA-GAAACAGCGCGCCCGACAGC 765
Db 140935  GCAAGTTACAGCTCCCTCGAGCTTTGAGAGACCGGAGAAATCAGCGCGCTCCACGTT 140994

QY 766  GCAGAAATCTTTGAGTTGCTCAGCTCGA---CTCCTCAGACCCCTCGGTCCACAGTCC 822
Db 140995  GCAGAAATCTTTGAGTTCTCGCGCTGGAATTTTCTTCGGGAAGACTTTGGCTGTCAGTTC 141054

QY 823  CTTTGAATTGGCTAGAGGGGATCCACTGGCAATACTCGCCCCAGACCTTAAATTTGGA 882
Db 141055  TCTTGATCGGCTAGAGGACGGACTCCAGTGGCAGTCTGCTCCTCAGACCTTGAATTTGA 141113

QY 883  GCTTAAGGCCAAGGGAAGTGCTTTGGACCTCAGCAGCAGGCTTTTCTCTAGAGCAGCA 942
Db 141114  ACTCAAGGCCCGGAAAGAGCTTTTAGACTCTACAGCGCACACTTTTCTCCTGAGAGCAGCA 141173

```

```

QY 943  GCTGTGGGAGTGGAGCTGTTTGGCCAGTAGCCTTCAATCCCGTCTGTACTCTAAACCGGA 1002
Db 141174  GCTGTGGGAGTGGAGTGTGCTGCCCGCTCTCTTCAACCGATAGTCTCTCCACCGAGA 141233

QY 1003  ACTTGGGCTCTTGGCCCTCTGGGCTCTTAAACATTTCAAACGATAGCAATTTCACTGGT 1062
Db 141234  ACTTGGGCTCTTCACTCTGGGCTCTGTGACACTCAGAGCTTGGGCAATTTCAAGGTAGT 141293

QY 1063  ATCTATTTGCTGAACCTTCTTCTACCTGGAGTGTCTTCTTAGCTAGAGTCACTATCA 1122
Db 141294  TTCTATCTCTGAACCTTCTTCTACCTGGAGTGTCTTCTTCCAGTTAGAGCTGCGCTGCA 141353

QY 1123  GAACAGTGTATGGAATAGCGAGGTAGTGGCTTCCAGACACTAAACCCAGAGAGCAGCTG 1182
Db 141354  GAACAGATCTGAGGTGGCCACTTGTGGATTTCCGAGCGCTATCCCGAGAGAGCTGTTA 141413

QY 1183  CCTGAGAGAGACCAATTTGTCATCCCGACCGCTGAGTGCAGAACTCATTCGGGCTCGTG 1242
Db 141414  CTTTCTGAGGTGGTTGTGCA-CTCCAGCGCTGCGTGTGAGATGTGCGCAACTGCTG 141472

QY 1243  GCAGGATGTCCACTCTTTCTACGGAAGGCTTACCAAGAAATTCACCATCTTCGCATGAA 1302
Db 141473  GCGGAGGTGTCCAGCCCTCTCTACAGAAAGGCTTCCGGAATTCACCATTCGGAATGAA 141532

QY 1303  ACGGCTGGAATTCCTTCAACAGGCTAAACAGGGGCAAGATTTACCCACCCCTGACCCAGGA 1362
Db 141533  ACGGCTAGAAATTCCTTC-----AGAGCAAGATTTACCAACCCCTGACCAAGA 141580

QY 1363  TAATGGCTTACACAGCCTGGAGGAGGAACACAGCCTTCTCGGATGGATCAAAACACTG 1422
Db 141581  CAATGGCTTATCATAGCTTGGAGGAGGAGCACTGCTTCTCGGATGGACCGCAACATTTG 141640

QY 1423  CAGAGATAACCCACACAGTTTGTCTGCTGCTGGAGACATTCCTCGGAAACACCCAGGA 1482
Db 141641  CACAGATAAGGTAGCAGGCTATGTCCCTGCTGGAG-----CCAGCCCGCA 141688

QY 1483  ATCCACTGAAGAAAAATAGAAATTAATACTACAGAGGTTTCCACTTCTTGTGGAGGAAGA 1542
Db 141689  GCCCACTGAGAAAAAAGTAGAATTTGGTGGTTGAGGAAGTTTCAC-----A 141733

QY 1543  GAGCCCTCTTGGAGGCTGTCCATCTAGTGAGATACCTATGGAAGAGAGCCTGAGAGGG 1602
Db 141734  GAGCCCGCAGGGAAGCAGTCTGTCTGTGAGTTAAGTGTGCAAAAGAAATGTGAAGAGGA 141793

QY 1603  CCGAATAAGTGTAGTTCATTACTCATACCTAGAGGTGACCTCCCATTTCTGCGCAGACC 1662
Db 141794  CCACACTAATGTAAATGACTACTCAGACATAGGGGGAACCTTCTGTTTCTCCAGACC 141853

QY 1663  AGCTTGTAGTAAACAAACTGATAGATTAATATTTTGGGAGGTGCAATCCAGTGCACCTGGAAC 1722
Db 141854  AGCTTGTACCAACAAACTGATAGATTAATATTTTGGGAGAGGCTCCAGTGCATTTGGACAC 141913

QY 1723  AAGTTCTGATCCAGAGGAGGATGGGATGAGGAAGCTGAGGATGATGTTTGTATAG 1782
Db 141914  CAGCTCCGCTTCTGAAAGTGGAGATTGGGATGAGGAACCTGAGTATGATGTTTGTATAG 141973

QY 1783  TGATAGCTCACTGTCAGACTCAGACCTTGAACAGACCCCTGGAAGGCTTCACTTTTGAA 1842
Db 141974  TGATGGCTCCCTGTCTGAAATCAGACAGGGAACAAAGATTCGAGAGGCTTCACTTTTGAA 142033

QY 1843  CTCCTTCTGCAAGTGTAGATCCTTATAATCCCGAAGCTTTTACAGCAACAAATTCAGACTGC 1902
Db 142034  CTCCTTCTACAGTGTAGATCCTTATAACACCCCAACACTTTACAGCCACGATTCAGACTGC 142093

QY 1903  TGCCAGAAATTTGCTCGAAGAGCCTTCTGATTCAGAGAGGATTTGTCTGGCAGTCTGA 1962
Db 142094  TGCCAGAAATTTGCCCCCGAGAGCCCATCTGATTCAGAGAGGCTCTGGTCTGGCAACCCGGA 142153

QY 1963  TCTAGAGAAATTCCTCCAGCTGTGGAAGCCTTCTTGAGACCCCTGAGCATAGTTCTGGGGA 2022
Db 142154  TGTAGGGA---GGTCTCAGGGCACACTCTCTTCCGAGACCCCTGACAGAGTTCTGGGGA 142210

```

QY 2033 GGAAGATGACGGGAATCTAGTGCAGATGAAGACGAGAGTCTCAAACTGTGGAACTCAT 2082  
 Db 142211 GGAGGATGACTGGGAACCTAGTGCAGAGGAAGCAGAAAATCATAGATTGTGGAACTCTCT 142270  
 QY 2083 CTGTAAATCTGATGACCCCTACACCCCTTTAAATTTTAAAGGCTCTTTTCAAAATCAGG 2142  
 Db 142271 CTGTAAATCTGAGGACCCCTACACCCCTTTAAATTTTAAAGGCTCTTTTCCAAAGCTTAGG 142330  
 QY 2143 GGAATAATCAGAAAGGCTGCTGCTGACTCAAGACCCCATCTGAGTCCATTTGTGGCCATTTTC 2202  
 Db 142331 GAAGAGCTGGAAGAGGCTGTGAGGACTCGAAGGCATCTCTGCGGCGCAGTGTCTTCTGTC 142390  
 QY 2203 TGAGTGCACACCTTACTTTCTTGAAGTGC-----AGCTGTGGGAGGACCAAGAAAGT 2257  
 Db 142391 TGGGCATCAGACCTTACTTTCTTGAAGGCAAGAACTTAAATTAGAGACCAAGAGA 142450  
 QY 2258 GAATGTCAGACTCGGTACAGCGTGACCTTTCTTTCGAGGAGACACACACATGTCAAA 2317  
 Db 142451 TAATTTGAGGCTGTGGGCTGGGTGAGGCTCTTTTGTGGAAGATACACCCATATCAAG 142510  
 QY 2318 AGAAAAAGGTAACTCTCTTGAAGAGTTACTGAGTATTATATAAGTGGTGATGAGGAT 2377  
 Db 142511 AGAAAAAGATAACTCTCTTGAAGAGTTACTGAGTATTATATAAGTGGTGATGAGGAT 142570  
 QY 2378 CGCAAGACCATGGGAAGATTTGCAAGGATGATGACGTTCCAGAAACGAATTCAA 2437  
 Db 142571 CGCAAGGATTTATGGGAAGATTTGCAAGGATGATGACGTTCCAGAAACGAATTCAG 142630  
 QY 2438 GAACAGAGATGCTATTGGATATGCTTGTGACATTTGAACACAGAGAAAGATGTTAAT 2497  
 Db 142631 GAACAGAGATGCTATTGGATATGCTTGTGACATTTGAACACAGAGAAAGATGTTAAT 142686  
 QY 2498 AGACTCCAGGGAACATGCTTCAAGAGACTTAATGTTCTCAAGCAATGTTGAGTTG-GCAG 2556  
 Db 142687 AGACTGCA-AGACCTGTTTCAAGAGACTTCTGTTGTACAGCTGTCTAGCAGAGGGA 142745  
 QY 2557 CCTGTAGCTCTAGTACATACATCTTCTTACCTGAGAGGTGCTTTTAAACCAAAAT 2616  
 Db 142746 GCAGCTGCAGCCCTCACCACTCCATCTCTTACTTGAAGAGTCTCTCTTATAGCAAAAC 142805  
 QY 2617 CTGGCAGCTGCTCTTGCATTTTCTTTTGTAGGAAAGTAACTTGGATCTAGTTTA 2676  
 Db 142806 ACTGCGAGCTGCTCTTGCATTTTCTTTTAAAGAAACGACTTGTCTTCTAGTGTAGTTT 142865  
 QY 2677 ATTTTCTTTTGTCAACATATCCCACTCAGAAACATTCAGGTTTGAAGCCAGCCCTGATA 2736  
 Db 142866 AGTTATTTTGGATATGTGCTCA-----TTAGAAACACACACAACTCTAATA 142914  
 QY 2737 ATGAAGATGAACATAGTGTGATTTCTAATCTCCCTTTTGTAGTTAGTTAGTGTGCTT 2796  
 Db 142915 ATGAAGAT-----CTCTAATCTTAACTCTTTCTTTCTATATAGTGTGATGTGGGT 142968  
 QY 2797 TTAATGTCCTTT 2809  
 Db 142969 TTTTCTTTTCTTT 142981  
 RESULT 14  
 AC094236/c  
 LOCUS AC094236 240836 bp DNA linear HTG 14-NOV-2002  
 DEFINITION Rattus norvegicus clone CH230-3B24, \*\*\* SEQUENCING IN PROGRESS \*\*\*,  
 4 unordered pieces.  
 ACCESSION AC094236  
 VERSION AC094236.5 GI:24958472  
 KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ENRICHED.  
 SOURCE Rattus norvegicus (Norway rat)  
 ORGANISM Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.  
 REFERENCE 1 (bases 1 to 240836)  
 AUTHORS Muzny,D.Marie., Metzker,M.Lee., Abramson,S., Adams,C., Alder,J.,  
 Allen,C., Allen,H., Alibrooks,S., Amin,A., Anguiano,D.,

Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,  
 Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,  
 Blawie,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,  
 Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,  
 Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,  
 Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,  
 Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,  
 Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,  
 Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,  
 Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,  
 Egan,A., Escotto,M., Eugene,C., Evans,C.A., Faller,T., Fan,G.,  
 Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,  
 Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,  
 Gehrgeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,  
 Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,  
 Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,  
 Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,  
 Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,  
 Jackson,B., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,  
 Karpachy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,  
 Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,  
 Liu,J., Liu,Y., Liu,Y., London,P., Longacre,S., Lopez,J.,  
 Lorensuewa,L., Loulesged,H., Lozado,R.J., Lu,X., Ma,J.,  
 Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,  
 Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,  
 Mahoney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,  
 Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,  
 Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,  
 Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,  
 Nwokeneme,O., Okwono,G., Olarnpungoon,A., Pal,S., Parks,K.,  
 Pasternak,S., Paul,H., Perez,A., Perez,L., Primus,E., Pu,L.-L.,  
 Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,  
 Puzo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,  
 Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,  
 Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,  
 Sanders,W., Savary,G., Scherer,S., Scott,G., Shadman,S., Shen,H.,  
 Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,  
 Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,  
 Steimle,M., Strong,R., Sutton,A., Svatek,A., Taber,P., Taylor,C.,  
 Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,  
 Valas,R., Vera,V., Villaseana,D., Waldron,L., Walker,B., Wang,J.,  
 Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,P.,  
 Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,  
 Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,  
 Yu,F., Zhang,J., Zhou,X., Zhao,S., Dunn,D., von  
 Niederhausen,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,  
 Weinstock,G. and Gibbs,R.A.  
 Direct Submission  
 Unpublished  
 2 (bases 1 to 240836)  
 Worley,K.C.  
 Direct Submission  
 Submitted (17-SEP-2001) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 240836)  
 Rat Genome Sequencing Consortium.  
 Direct Submission  
 Submitted (14-NOV-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 On Nov 14, 2002 this sequence version replaced gi:22772949.  
 The sequence in this assembly is a combination of BAC based reads  
 and whole genome shotgun sequencing reads assembled using Atlas  
 (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described  
 in the feature table below represents a scaffold in the Atlas  
 assembly (a 'contig-scaffold'). Within each contig-scaffold,  
 individual sequence contigs are ordered and oriented, and separated  
 by sized gaps filled with Ns to the estimated size. The sequence  
 may extend beyond the ends of the clone and there may be sequence  
 contigs within a contig-scaffold that consist entirely of whole  
 genome shotgun sequence reads. Both end sequences and whole genome  
 shotgun sequence only contigs will be indicated in the feature

table.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: GAGF  
Center clone name: CH230-3B24  
----- Summary Statistics

Assembly program: Phrap; version 0.990329  
Consensus quality: 194130 bases at least Q40  
Consensus quality: 198891 bases at least Q30  
Consensus quality: 202078 bases at least Q20  
Estimated insert size: 194853; sum-of-contigs estimation  
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

-----  
\* NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 4 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 184581: contig of 184581 bp in length  
\* 184582 184681: gap of unknown length  
\* 184682 238285: contig of 53604 bp in length  
\* 238286 238385: gap of unknown length  
\* 238386 239477: contig of 1092 bp in length  
\* 239478 239577: gap of unknown length  
\* 239578 240836: contig of 1259 bp in length.

# FEATURES

source

1..240836  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clone="CH230-3B24"  
misc\_feature  
1984..3036  
/note="wgs\_contig"

# ORIGIN

Query Match 33.7%; Score 991.4; DB 2; Length 240836;  
Best Local Similarity 72.3%; Pred. No. 1.9e-256;  
Matches 1494; Conservative 0; Mismatches 496; Indels 75; Gaps 13;  
  
QY 754 CGCCCCCAGCGCAGAAATCTTTGAGTTTCGCTGCGAGTCGA---CTCCTCAGACCCCTC 810  
DB 174258 CGCTCCAGTTGCAGAAATCTTTGAGTTTCGCTGCGAGTCGAATTTTCTTGGAGACTT 174199  
  
QY 811 GGTACACAGTCCCTTGTGCTAGAGAGGGGATCCACTGGCAATFACCTGCCCCCAGA 870  
DB 174198 GGCTGTGCTTCTTGTGCTGAGAGCGGACTCCAGTGGCAGTGTCTGCTCAGA 174139  
  
QY 871 CCTAAATTTGAGACTTAAGGCCAAGGAAGTCTTTTGGACCTCGAGCACAGGCTTTTCT 930  
DB 174138 CTTGGAATTTGAACTCAGAGGCCCCAGAAAGAGCTTTAGACTCTAGAGCGCACACTTTCT 174080  
  
QY 931 CTTAGAGCAGCAGCTGTGGGAGTGGAGCTGTGCCCAGTAGCCCTTCAATCCCGTCTGTA 990  
DB 174079 CATGAGCAGCAGCTGTGGGAGTGGAGTTGTGCTGCCCCCTCTCTTCAAACCGACTAGT 174020  
  
QY 991 CTCTAACCGGAACTTGGCTCTTGGCCCTCTGGGCTCTAAACATTTCAACGATAGACA 1050  
DB 174019 CTCCCAACGAAGAACTTGGCTCTTCACTCTGCGGCTCTGAGCACTCAGAGCTTGGGCAA 173960  
  
QY 1051 TTTTCAGTGTGATCTCTATTTCTGAAACCTTCTTACCTGGAGTCTGTTTCTTAGGCTAGA 1110  
DB 173959 TTTCAAGGTAGTTTCTATCTCTGAAACCTTCTTACCTGGAGTCTGCTTCCCGACTAGA 173900  
  
QY 1111 AGTCAGCTATCAGAACAGGTGATGGAATAGCGAGGTAGTCGGCTTCCAGACACTAACCCC 1170

Db	173899	GCTCGCTGTGCTGAGAACAGCATCTGAGGTGGCCAACTTTGTGATTTCCGAGCGCTATCCCC	173840
QY	1171	AGAGAGCAGCTGCTGAGAGGAGCATTTGTCTATCCCGACCGCTGAGTGCAGAACTCAT	1230
Db	173839	AGAGAGCTGTGTACCTTTCTTGAAGGTGTGTGTCA-CTCAGCGCTGCTGCTGAGATGTC	173781
QY	1231	TCCGCGCTCGTGGCAGGATGCTCCACCTCTTCTACGGAAGGCTTACAGAAATTCACCA	1290
Db	173780	GGCAACTGCTGGCGGAGGTGTCCAGCCCTCTCTACAGAGGCTTCCGGAATTCACCA	173721
QY	1291	TCCTTCGATGAAACCGCTGGAATTCCTTCAACAGGCTAAACAAGGGGCAAGATTACCCAC	1350
Db	173720	CATTCCGATGAAACCGCTAGAAATTCCTTC-----AGAGGCAAGAGTTACCAAC	173673
QY	1351	CCCTGACCGAGTAATGCTTACACAGCCTGGAGGAGAACACAGCTTCTCCCGATGGA	1410
Db	173672	CCCTGACCGAGCAATGGCTATCATGCTTGGAGGAGGAGCCTTCTTCCCGATGGA	173613
QY	1411	TCCAAACACTGACAGATAACCCCAACACAGTTTGTTCCTGCTGCTGAGACATTCCTGG	1470
Db	173612	CCCGCACTTGCACAGTAAGGTAGCACAGGCTATGTCCCTGCTGGAG-----	173563
QY	1471	AAACCCAGGAATCCACTGAAAGAAATAAGAAATTAATACTACAGAGTTCCACTTGC	1530
Db	173562	--CCAGCCCGCAGCCCACTGAGAAAAAAGTAGAATTGCTGTTGAGGAAGTTTCAC----	173509
QY	1531	TTTGGAGAGAGAGAGCCCTTCTGAGGCTGCTCCATCTAGTCAGATACCTATGGAAGA	1590
Db	173508	-----AGAGCCCCCAGGGAAGAGTCTCTCTGTGAGTTAAGTTGCAAAAGA	173460
QY	1591	GCCTGGAGAGGCGCGAATAAGTGTAGTTGATTACTCATCTACAGAGTGCCTTCCCAT	1650
Db	173459	ATGTGAAGAGGACCACTAATGTAAATTGACTACTCAGACATAGGGGGAACCTTCTGT	173400
QY	1651	TTCTGCCAGACCACTTGTAGTAACAACTGATAGATTATTTTGGAGGTGCATCCAG	1710
Db	173399	TTCTCCAGACCACTTGTATACCAACAACTGATAGATTATTTTGGAGGAGCTCCAG	173340
QY	1711	TGACCTGGAAACAAAGTTCTGATCCAGAGGTGAGGATTTGGATGAGGAAGCTGAGGATGA	1770
Db	173339	TGACTTGGACACCACTGCTCCGTTTCTGAAAGTGAGGATTTGGATGAGGAACCTGAGTATGA	173280
QY	1771	TGTTTGTGATAGTATAGCTCACTGTGAGACTCAGACTTGAAACAAGCCCTGAAGGCT	1830
Db	173279	TGGTTTGTGATAGTATGCTCCCTGTCTGAATCAGACAGGGAACAAGACTCCGAAGGCT	173220
QY	1831	TCACCTTTGGAACTCTTTTCTGAGTGTAGATCTTATATATCCCGCAGAACTTTACAGCAAC	1890
Db	173219	TCACCTTTGGAACTCTTTTCTACAGTGTAGATCTTACAAACCCCAACACTTTACAGCCAC	173160
QY	1891	AAATTCAGACTGCTGCCAGAAATTTCTTGAAGGCTTCTGATTCAGAGAGGATTTGTCT	1950
Db	173159	GATTCAGACTGCTGCCAGAAATTTGCCCAAGAGACCCATCTGANTCAGAGAGTCTGTGTC	173100
QY	1951	TGGCAAGTCTGATCTAGAGAAATTCCTCCAGTCTGGAAGCCTTCTGAGACCCCTGAGCA	2010
Db	173099	TGGCAACCCGATGTAGGA---GGTCTCAGGCCACACTCTTCCGAGAGCCCTGACCA	173043
QY	2011	TAGTTCTGGGAGGAAAGATGATCTGGGAATCTAGTGTGAGATGAAGCAGAGACTCTCAAACT	2070
Db	173042	GAGTTCTGGGAGGAGGATGATCTGGGAACCTTAGTGTGAGAGGAAGCAGAAAAATCATAGATT	172983
QY	2071	GTGGAATCTATTCTGTAAATTTCTGATGATGATCAACCCCTTTAAATTTTAAAGGCTCTTTT	2130
Db	172982	GTGGAATCTCTCTGTAAATTTCTGAGGACCCCTACAACTTTTAAATTTTAAAGGCTCTTTT	172923
QY	2131	TCAAACATCAGGGGAAATGAGAAAGGCTGTGCTGACTCAAAAGACCCCATCTAGTCCAT	2190
Db	172922	CCAAACGTTAGGGAAGAGCTGGAAGGCTGTGAGGACTGGAAGGCTCTCTGCGGGCCAC	172863
QY	2191	TGTGGCAATTTCTGAGTGTACACCTTACTTCTTGTGTAAGTGC-----AGCTGTGGGG	2245



```
Db 172862 AGTGGTCTTGTCTGGGCATCAGACTTACTTCTTTGTAAGGCACAGAAACTTAATTAGA 172803
Qy 2246 AGCCAGAAAGTGAATGTCCAGACTCGGTACAGCGTGACGTTCTTCTTGGAGGAAGACAC 2305
Db 172802 GAGCCAGAGAGATAATTGTAGGGCTGTGGGCTGGGTGAGCTCTTTTGGTGAAGATAC 172743
Qy 2306 ACACATGTCAAAAGAAAAGGTAACTTCTTCTTGAAGAGTGTACTGAGTATATATAAGT 2365
Db 172742 ACCCATATCAAGAGAAAAGATAAACCCTTCTTGAAGAGTGTACTGAGTATATATAAGT 172683
Qy 2366 GGTGATCGAGATCCCAAGGACCAATGGAGAGATTTCCAGGGAGTGCATGAGTCCAG 2425
Db 172682 GGTGATGAGGCTCCCAAGGATTAATGGAGAGATTTGCAAGGAGTGCATGAGTCCAG 172623
Qy 2426 AAACGAATTCAAGAAAACAGAGATGCTATTGGATATTGCTTGACATTTGACACACAGAGAA 2485
Db 172622 AAACGAATTCAAGAAAACAGAGATTGCCATTGGTTACTGCTTGAGCTTTGAGCAC- ---AA 172567
Qy 2486 AGAATGTTTAATAGACTCCAGGGAACATGCTTCAAGAGACTTAATGTTCTCAAGCAATGT 2545
Db 172566 AGAGTGTTTAGTAGACTGC-GAGACCGTGTGTTCAAGAGACTTCTGTTGTACAGCAGTGC 172508
Qy 2546 TGAGTTG-GCAGCTGTAGTCTAGCTAGCATACACTACCTCTTACCTGAGAGTGTCTT 2604
Db 172507 TAAACAGAGGAGCAGCCTGCAGCCCTCACCCACTCCATCTCTTACTTTGAGAGGGTTCTC 172448
Qy 2605 TTAAGAAACAAATCTTGGCAGCTGCTTGTGACATTTTCTTGTGAGGAAATGTAACCTT 2664
Db 172447 TTATAGACAAACACTGGCAGCTGCTTGTGACATGGGTTTAAGAAACGACTTGGTTCTA 172388
Qy 2665 GGATCTAGTTAAATTTTTTTTTTGGCAACATATCCCACTCGAAGAACATTCAGGTTTGAAG 2724
Db 172387 GTGATGATGTTAGTTATTTTGGATAATGTGTCTCA-----TTAGAAACACAC 172339
Qy 2725 CCAGCCCTGATAATGAGGAGTACTAGTGTGATTTCTTAATCTCCCTTTTGTGATTTAG 2784
Db 172338 ACAACTCTAAATGAGGGAT-----CTCTAATCTTTAATCTTCTTTTCTTATATAG 172285
Qy 2785 TTGGATGTGCTTTTAAATGTCTTT 2809
Db 172284 CTGGATGTGGTTTTTTTTTTTTTTTT 172260

RESULT 15
BD148615 772 bp DNA linear PAT 17-JAN-2003
LOCUS Primer for synthesizing full-length cDNA and use thereof.
DEFINITION
BD148615
ACCESSION BD148615.1 GI:27854373
VERSION JP 2002191363-A/3458.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
Ota,T., Teogai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J.,
Iehi,S., Sugiyama,T., Wakamatsu,A., Nagai,K. and Otauka,T.
Primer for synthesizing full-length cDNA and use thereof
Patent: JP 2002191363-A 3458 09-JUL-2002;
HELIX RESEARCH INSTITUTE
OS Homo sapiens (human)
PN JP 2002191363-A/3458
PD 09-JUL-2002
PF 28-JUL-2000 JP 2000280990
PI TOSHIO OTA,TAKAO ISOAI,TETSUO NISHIKAWA,KOJI HAYASHI,KAORU
PI SAITO,
PI JUNICHI YAMAMOTO,SHIZUKO ISHII,TOMOYASU SUGIYAMA,AI WAKAMATSU,
PI KEIICHI NAGAI,TETSUJI OTSUKI
PC
C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N5/ PC
10,
PC C12P21/02,C12Q1/68/C12P21/08,G06F17/30,C12N15/00,C12N5/00 CC
Primer for synthesizing full-length cDNA and use thereof FH Key
```

```
FEATURES
source
FT source Location/Qualifiers
FT 1..772
/organism="Homo sapiens (human)"
Location/Qualifiers
1..772
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 25.1%; Score 738; DB 6; Length 772;
Best Local Similarity 98.8%; Pred. No. 4.2e-188; Indels 1; Gaps 1;
Matches 752; Conservative 0; Mismatches 8;
Qy 1 ATTTTGGGCTTCGCTTCCACCGCACCGCGGCTACCCAGTCTTCCGATATCGGTTG 60
Db 1 ATTTTGGGCTTCGCTTCCACCGCACCGCGGCTACCCAGTCTTCCGATATCGGTTG 60
Qy 61 CTCAGGGGCTTTTCAACCCCTCTGTCTAGTTCGGAACCATCGCCGAGGCGGTGGGGGACT 120
Db 61 CTCAGGGGCTTTTCAACCCCTCTGTCTAGTTCGGAACCATCGCCGAGGCGGTGGGGGACT 120
Qy 121 CCTATCCATGCTGTTGAAGGCTCGAGCCGACTAGGGAACCTCTCTCCGCCAGGATGGA 180
Db 121 CCTATCCATGCTGTTGAAGGCTCGAGCCGACTAGGGAACCTCTCTCCGCCAGGATGGA 180
Qy 181 AGTCGATCAGTCGCGCCCTATTTCGCGGGGCTGTTCTTCCCTGTGTTCTGCCGCCGCTG 240
Db 181 AGTCGATCAGTCGCGCCCTATTTCGCGGGGCTGTTCTTCCCTGTGTTCTGCCGCCGCTG 240
Qy 241 CCGCATTCGCTGCTCTGTGGCTTTTCTGCTGGCTTTCGCTGGAAGATCGGCTCGAGCGGAC 300
Db 241 CCGCATTCGCTGCTCTGTGGCTTTTCTGCTGGCTTTCGCTGGAAGATCGGCTCGAGCGGAC 300
Qy 301 GCCACCGCTGGGAAGCCGAGACTCTGTAGGCTTCTCCGAATCCGCTCGACTCTCCAGC 360
Db 301 GCCACCGCTGGGAAGCCGAGACTCTGTAGGCTTCTCCGAATCCGCTCGACTCTCCAGC 360
Qy 361 CGCTGAGCGCGCGGCTTCTACCTGAGAGACTGTCAAGAAAAGAGATGAGAGCGGGGAC 420
Db 361 CGCTGAGCGCGCGGCTTCTACCTGAGAGACTGTCAAGAAAAGAGATGAGAGCGGGGAC 420
Qy 421 AGCGGATTCGCGAAACGGCTTGGCCCTCGGGCGGGCTTCCGCTTCTGGCCACCTTTT 480
Db 421 AGCGGATTCGCGAAACGGCTTGGCCCTCGGGCGGGCTTCCGCTTCTGGCCACCTTTT 480
Qy 481 CCCTCGCGGATTCGCAAGCAGGCTTCTTAAGTTCCGAGCGCTTCTGGCCCGGAAACTC 540
Db 481 CCCTCGCGGATTCGCAAGCAGGCTTCTTAAGTTCCGAGCGCTTCTGGCCCGGAAACTC 540
Qy 541 CGGGAACCCCACTGCTTCTCTGCCAGCCGAGACTCGGGTCACTTACTTGGACGAA 600
Db 541 CGGGAACCCCACTGCTTCTCTGCCAGCCGAGACTCGGGTCACTTACTTGGACGAA 600
Qy 601 ACTGCTCTCCAGCTCTTTCGCGGCTTCCCGGATTCCTTCAGAGTGTCTAATTTGGAG 660
Db 601 ACTGCTCTCCAGCTCTTTCGCGGCTTCCCGGATTCCTTCAGAGTGTCTAATTTGGAG 660
Qy 661 CCAACTTTTTCGGTGGAAATGTTTCCGACCAAGATGGTAGATTTTGTGGAGTCTAC-AGCG 719
Db 661 CCAACTTTTTCGGTGGAAATGTTTCCGACCAAGATGGTAGATTTTGTGGAGTCTACAGCG 720
Qy 720 CCCTGAGAGCCCTGAAGGGAGCGGAGAAACACAGCCGCCCC 760
Db 721 CCCTGAGAGCCCTGAAGGGAGCGGAGAAACCAAGCCGCCCC 761
```

Search completed: September 16, 2005, 01:57:34  
Job time : 12271.5 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 15, 2005, 08:34:33 ; Search time 22770.5 Seconds  
(without alignments)  
11635.782 Million cell updates/sec

Title: US-10-650-482-3  
Perfect score: 5468  
Sequence: 1 cggctcctcgctcgccgtg.....attaaaaaaaaaaaaa 5468

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 100 summaries

Database :

GenEmbl.\*  
1: gb\_ba.\*  
2: gb\_htg.\*  
3: gb\_in.\*  
4: gb\_om.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pl.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_sts.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: gb\_vi.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5234.2	95.7	5367	10	BC058078 Mus muscu
2	3115	57.0	3704	10	BC068897 Mus muscu
3	3100	56.7	215781	10	AC107842 Mus muscu
4	2412.4	44.1	212261	2	AC119774 Rattus no
C 5	2188.8	40.0	240836	2	AC094236 Rattus no
C 6	1689.2	30.9	200490	2	AC128396 Rattus no
C 7	1689.2	30.9	227054	2	AC098956 Rattus no
8	1689.2	30.9	248458	2	AC105827 Rattus no
9	1320.8	24.2	2942	6	BD160691 Primer fo
10	1320.8	24.2	2942	6	AX884059 Sequence
11	1320.8	24.2	2942	9	AK027650 Homo sapi
12	1320.8	24.2	5264	9	BC065280 Homo sapi
C 13	1124.8	20.6	167079	9	AL606489 Human DNA
C 14	1124.8	20.6	328863	2	AL627202 Rattus no
C 15	671	12.3	166233	2	AC048369 Homo sapi
16	489.4	9.0	257938	2	AC124919 Rattus no
17	353	6.5	263621	2	AC095784 Rattus no
C 18	262.6	4.8	212115	2	AC116701 Mus muscu
19	239.2	4.4	772	6	BD148615 Primer fo

93 96.2 1.8 316339 2 AC095802 AC095802 Rattus no  
 94 96 1.8 119664 10 AC091537 AC091537 Rattus no  
 95 96 1.8 136416 2 AC145754 AC145754 Rattus no  
 c 96 96 1.8 239665 2 AC127136 AC127136 Rattus no  
 97 96 1.8 251882 2 AC096474 AC096474 Rattus no  
 98 96 1.8 262320 2 AC135741 AC135741 Rattus no  
 99 96 1.8 274522 2 AC110095 AC110095 Rattus no  
 100 95.8 1.8 240239 2 AC123154 AC123154 Rattus no

## ALIGNMENTS

RESULT 1  
 BC058078  
 LOCUS  
 DEFINITION  
 15b, mRNA (cdna clone MGC:66517 IMAGE:6830215), complete cds.

ACCESSION  
 BC058078  
 VERSION  
 MGC.  
 KEYWORDS  
 BC058078.1 GI:34785997  
 SOURCE  
 Mus musculus (house mouse)

ORGANISM  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE  
 AUTHORS  
 1 (bases 1 to 5367)  
 Straube, R.D., Collins, F.S., Wagner, L.H., Derge, J.G.,  
 Klausner, R.D., Collins, F.S., Wagner, L.H., Shenmen, C.M., Schuler, G.D.,  
 Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,  
 Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,  
 Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,  
 Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,  
 Scheetz, T.E., Brownstein, M.J., Uedin, T.B., Toehiyuki, S.,  
 Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,  
 Abramson, R.D., Mullany, S.J., Bosak, S.A., McEwan, P.J.,  
 McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,  
 Woxley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,  
 Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,  
 Pahey, J., Helton, E., Kettner, M., Madan, A., Rodriguez, S.,  
 Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,  
 Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,  
 Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,  
 Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalusz, D.E.,  
 Schnerch, A., Schein, J.E., Jones, S.J., and Marra, M.A.

Generation and initial analysis of more than 15,000 full-length

human and mouse cDNA sequences

Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

22388257

12477932

2 (bases 1 to 5367)

Straube, R.

Direct Submission

Submitted (08-SEP-2003) National Institutes of Health, Mammalian

Gene Collection (MGC), Cancer Genomics Office, National Cancer

Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,

USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk

Email: [cgapsb-remail.nih.gov](mailto:cgapsb-remail.nih.gov)

Tissue Procurement: Dr. Jim Lin, University of Iowa

cDNA Library Preparation: M. Bento Soares, University of Iowa

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Genome Sequence Centre,

BC Cancer Agency, Vancouver, BC, Canada

[info@bcgsc.bc.ca](mailto:info@bcgsc.bc.ca)

Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,  
 Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,  
 Leticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo  
 Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven  
 Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline  
 Schein, Duane Smalusz, Michael Teal, Lorraine Spence, Jeff Stott,  
 Michael Thorne, Miranada Teal, Natasja van den Bosch, Jill Vardy,  
 George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found  
 through the I.M.A.G.E. Consortium/LINL at: <http://image.llnl.gov>  
 Series: IRAC Plate: 126 Row: n Column: 5  
 This clone was selected for full length sequencing because it  
 passed the following selection criteria: matched mRNA gi: 27597066.

## FEATURES

## source

1. 5367  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6"  
 /db\_xref="taxon:10090"  
 /clone="MGC:66517 IMAGE:6830215"  
 /tissue\_type="Brain, enriched mouse brain 12.5dp"  
 /clone\_lib="NIH\_BMAP\_F00"  
 /lab\_host="DH10E"  
 /notes="vector: pYX-ASC"  
 1. 5367  
 /gene="Ppplr15b"  
 /db\_xref="LocusID:108954"  
 /db\_xref="MGI:2444211"  
 365..1591  
 /gene="Ppplr15b"  
 /codon\_start=1  
 /product="Ppplr15b protein"  
 /protein\_id="AAH58078.1"  
 /db\_xref="GI:34785998"  
 /db\_xref="LocusID:108954"  
 /db\_xref="MGI:2444211"  
 /translation="METGTHARKRPGLSGWFLPFLRRSHACSEFPFPPSRQNP  
 LNSALPERTRYTKLSQLLALPSLQKLLWSQLFGLLPTRMWDLFAASYSALRA  
 GRSERAAFTQKLSLSDSDSLVSSLDWLEGLQWCSSSDLKLAQERA  
 LDSAAPTFLLEQWLWGLVPSLIQAGLVSHRELDSSSGPLSVQSLGNFKVSYLLN  
 PSYLDVLPGILGRCOSSAGGQFVGFRITLTPESCYSLEDGCHPQIARAENSATWARRC  
 PPISTEGLEPIHILMKLEFLQANKGOELPTPDQNGVHSLREHNLLRMDPOHCTD  
 NPAQAVSPADRPEKTEKPELVIQEVSQFQSSSLFCELPVEKECEDHTNATDLS  
 RGSLEPVPSTRPVVATN"

## ORIGIN

Query Match 95.7%; Score 5234.2; DB 10; Length 5367;  
 Best Local Similarity 99.5%; Pred. No. 0;  
 Matches 5345; Conservative 0; Mismatches 18; Indels 10; Gaps 9;  
 Qy 98 CGCCTTGGCGTTTCGCGCCTCCGTCGGCGGCACCGGAAACCGCGCCTCGCTCGCGC 157  
 Db 1 CGCCTTGGCGTTTCGCGCCTCCGTCGGCGGCACCGGAAACCGCGCCTCGCTCGCGC 60  
 Qy 158 CGCGCGGAGGAGGGTCTTCTATGTGGAGCGATCTCACACGGCCTAGGACGCTC 217  
 Db 61 CGCGCGGAGGAGGGTCTTCTATGTGGAGCGATCTCACACGGCCTAGGACGCTC 120  
 Qy 218 CTTCCCTAGCGGGATGGACCTAACCGCGGTTCGCCACCGTTCGGCGGCCTCTGGGCG 277  
 Db 121 CTTCCCTAGCGGGATGGACCTAACCGCGGTTCGCCACCGTTCGGCGGCCTCTGGGCG 180  
 Qy 278 TCCGGTGCAGCACTCGTTGCGGAGCGCGCTCTCTGGGCTCTCTGCGCGCGCGGA 337  
 Db 181 TCCGGTGCAGCACTCGTTGCGGAGCGCGCTCTCTGGGCTCTCTGCGCGCGCGGA 240  
 Qy 338 ATCGGACTGCAGTACCACTCCGTCGGCTGGGCAAGGCGGAGACTGTGTAGACCTCGGATC 397  
 Db 241 ATCGGACTGCAGTACCACTCCGTCGGCTGGGCAAGGCGGAGACTGTGTAGACCTCGGATC 300  
 Qy 398 CAGCTCGGCTGACGGCGCTGAGCTCTGTCTCTCTCTGTCTGAGAGAGCCGCCAAGGAAA 457  
 Db 301 CAGCTCGGCTGACGGCGCTGAGCTCTGTCTCTCTCTGTCTGAGAGAGCCGCCAAGGAAA 360  
 Qy 458 GGAGATGGAGACAGGAACGACAGGGCCCGGAAAGCGCCTCGGCTCGGCTCTG 517  
 Db 361 GGAGATGGAGACAGGAACGACAGGGCCCGGAAAGCGCCTCGGCTCGGCTCTG 420  
 Qy 518 GTTCGGCTCGGCTCTCTCTCGGAGATCGACGCTGCTCTTCGAGTTCGCGCGCCTTC 577

Db 421 GTTCCGGCTGCCCTTCCTTCGGGGATCGCAGCGCTGCTCTTCGGAGTTCGCCGCGCTTC 480  
Qy 578 CTCTCGACAAAATCCCGGAACTCCGCTCTGCCCGAGCGTCGACACGAGTACTGGACCAA 637  
Db 481 CTCTCGACAAAATCCCGGAACTCCGCTCTGCCCGAGCGTCGACACGAGTACTGGACCAA 540  
Qy 638 ATTGCTTTCTCAGCTCTTGGCTGCTCCCTAGCTATTTCCAGAAAGTGTCTGCTTTGGAG 697  
Db 541 ATTGCTTTCTCAGCTCTTGGCTGCTCCCTAGCTATTTCCAGAAAGTGTCTGCTTTGGAG 600  
Qy 698 CCAGCTTTCCGGGGCTGATTTCTACAGATGCTAGATTTTGGCCGAAGTTACAGCGC 757  
Db 601 CCAGCTTTCCGGGGCTGATTTCTACAGATGCTAGATTTTGGCCGAAGTTACAGCGC 660  
Qy 758 CCTGAGAGCTTCGAGAGGACGGAGGAATCTGACGCTCCACCGTGCAGAAAGTCTCTGAG 817  
Db 661 CCTGAGAGCTTTGAGAGGACGGAGGAATCTGCGCTCCCAAGTGCAGAAAGTCTCTGAG 720  
Qy 818 TTACATGCGGCTGGACTCTTTCGGGAAGACTCGCGTCTGTCAGTACTCTTTCGATTTGGCTAG 877  
Db 721 TT-CACCTGCGGCTGGACTCTTTCGGAAGACTT--GGTTGTCAAGTCTCTTG-ATTGGCTAG 776  
Qy 878 AGAGGAGCTTCCAGTGGGAGTCTCGTCCCTCAGA-CTGGAAGTTAAACTCAAGGCCCAGG 936  
Db 777 AGAGGAGCTTCCAGTGGGAGTCTCGTCCCTCAGACCTGGAGTTAAAACTCAAGGCCCAGG 836  
Qy 937 AAAGAGCTTTAGACTCTGCAGCGCCCACTTTCTCTGAGCAGCAGCTGTGGGGAGTGG 996  
Db 837 AAAGAGCTTTAGACTCTGCAGCGCCCACTTTCTCTGAGCAGCAGCTGTGGGGAGTGG 896  
Qy 997 AGTTGCTGCCAGTAGCCTTCAAGCTGGTCTAGTCTCCCAAGCAGAACTTGACTCTTTCAT 1056  
Db 897 AGTTGCTGCCAGTAGCCTTCAAGCTGGTCTAGTCTCCCAAGCAGAACTTGACTCTTTCAT 956  
Qy 1057 CCTCTGGGCTCTGAGGCTTCAGAGCTTAGGTAATTTCAAGTAGTTTCTATCTCCTGA 1116  
Db 957 CCTCTGGGCTCTGAGGCTTCAGAGCTTAGGTAATTTCAAGTAGTTTCTATCTCCTGA 1016  
Qy 1117 ACCCTTCTACCTGGACTACCTTCCCAAGTTAGGGCTGCGCTGTCAGAGCAGCGCTGGAG 1176  
Db 1017 ACCCTTCTACCTGGACTACCTTCCCAAGTTAGGGCTGCGCTGTCAGAGCAGCGCTGGAG 1076  
Qy 1177 GTGGCCAGTTTGTGGGTTTCCGAACACTAAACCCAGAGAGTGTCTATCTTTCTGAAGATG 1236  
Db 1077 GTGGCCAGTTTGTGGGTTTCCGAACACTAAACCCAGAGAGTGTCTATCTTTCTGAAGATG 1136  
Qy 1237 GTTGTCACTCAGCGCTTGGGGCAGAGATGTCCGCAACCGCTGAGAGAGTGTCCGC 1296  
Db 1137 GTTGTCACTCAGCGCTTGGGGCAGAGATGTCCGCAACCGCTGAGAGAGTGTCCGC 1196  
Qy 1297 CTCTCTACAGAAAGGCTGCGGAAATCCACCACTTCGTATGAAACGGCTAGAAATTC 1356  
Db 1197 CTCTCTACAGAAAGGCTGCGGAAATCCACCACTTCGTATGAAACGGCTAGAAATTC 1256  
Qy 1357 TCCAGGCTTAAACAAAGGCAAGAGTTAACCACCCCTGACCAAGATAATGGCTATCATAGCC 1416  
Db 1257 TCCAGGCTTAAACAAAGGCAAGAGTTAACCACCCCTGACCAAGATAATGGCTATCATAGCC 1316  
Qy 1417 TGGAGGAGAACATAACCTTCTCGGATGAGCCCAACACATTGCAAGATAACCCAGCAC 1476  
Db 1317 TGGAGGAGAACATAACCTTCTCGGATGAGCCCAACACATTGCAAGATAACCCAGCAC 1376  
Qy 1477 AGGCGGTGTCCCTGCTGCAGACAGGCGGAGCCCACTGAGAAAAAACAGAAATTTGGTGA 1536  
Db 1377 AGGCGGTGTCCCTGCTGCAGACAGGCGGAGCCCACTGAGAAAAAACAGAAATTTGGTGA 1436  
Qy 1537 TTCAGAAGTTTTCACAGAGCCCCAGGGAAAGCAGTCTGTTTTGTGAATTAACCCGTGAAA 1596  
Db 1437 TTCAGAAGTTTTCACAGAGCCCCAGGGAAAGCAGTCTGTTTTGTGAATTAACCCGTGAAA 1496  
Qy 1597 AAGNATGTGAAGAGCACACTAATGCAACTGACCTCTCAGATAGAGGAGAGCCCTTC 1656  
Db 1497 AAGNATGTGAAGAGCACACTAATGCAACTGACCTCTCAGATAGAGGAGAGCCCTTC 1556

Qy 1657 CTGTTTTCTACGACCAAGTTTGTAGCAACAAACTGATAGATTATATTTTGGGAGGCGCC 1716  
Db 1557 CTGTTTTCTACGACCAAG-TTGTAGCAAACAACTGATAGATTATATTTTGGGAGGCGCC 1615  
Qy 1717 CAGTCACTTGGAGCCAGCTCTGATTTCTGAAGTGAAGTTGGGGGAGGAACCTTGAGG 1776  
Db 1616 CAGTCACTTGGAGCCAGCTCTGATTTCTGAAGTGAAGTTGGGGGAGGAACCTTGAGG 1675  
Qy 1777 ACCATGGCTTTGATAGCGATGGCTCCCTGTCTGAATCAGACGTGGAAACAGGACTCGGAAG 1836  
Db 1676 ACGATGGCTTTGATAGCGATGGCTCCCTGTCTGAATCAGACGTGGAAACAGGACTCGGAAG 1735  
Qy 1837 GCCTTCACTTTTGGAACTCTTTTCCACAGTGTAGATCTTTTCAAAACCCCAAAACTTTTACAG 1896  
Db 1736 GCCTTCACTTTTGGAACTCTTTTCCACAGTGTAGATCTTTTCAAAACCCCAAAACTTTTACAG 1795  
Qy 1897 CCACGATTCAGACGGCTGCCAGAAATTTGCCCCACAGACCCCATCAGATTTCAAGGACATCCT 1956  
Db 1796 CCACGATTCAGACGGCTGCCAGAAATTTGCCCCACAGACCCCATCAGATTTCAAGGACATCCT 1855  
Qy 1957 GGTCTGGCAGCTGTGTGTAGGAGCTGTCCAGGAGGACCCCTTCCGGAGACCCCGACC 2016  
Db 1856 GGTCTGGCAGCTGTGTGTAGGAGCTGTCCAGGAGGACCCCTTCCGGAGACCCCGACC 1915  
Qy 2017 ATAGTTCCGGGGAGGAAGATGACTGGGAACCCGAGTGCAGATGAAGCAGAGAACTTTAAAT 2076  
Db 1916 ATAGTTCCGGGGAGGAAGATGACTGGGAACCCGAGTGCAGATGAAGCAGAGAACTTTAAAT 1975  
Qy 2077 TGTGGAACCTTTTCTGTCATTTCTGAGGACCCCTTACAACTTTTAAATTTTAAAGCTCCTT 2136  
Db 1976 TGTGGAACCTTTTCTGTCATTTCTGAGGACCCCTTACAACTTTTAAATTTTAAAGCTCCTT 2035  
Qy 2137 TTCAAACCGTCAGGAGAAATTTGAAAGGCGCTCAGGACTCAAAGGCGCTTCTTCTGAGGTCA 2196  
Db 2036 TTCAAACCGTCAGGAGAAATTTGAAAGGCGCTCAGGACTCAAAGGCGCTTCTTCTGAGGTCA 2095  
Qy 2197 CAGTGCCTTCTCTGCGCATCATACCTTACTTTCTTGTAAAGCCAGCTGTTTGAAGAGCC 2256  
Db 2096 CAGTGCCTTCTCTGCGCATCATACCTTACTTTCTTGTAAAGCCAGCTGTTTGAAGAGCC 2155  
Qy 2257 AAGAAATTAATTTCTCAGGCTGTGGCTGGGTGAGGCTCTTGTCTGGAAGAAAGTACACC 2316  
Db 2156 AAGAAATTAATTTCTCAGGCTGTGGCTGGGTGAGGCTCTTGTCTGGAAGAAAGTACACC 2215  
Qy 2317 ATATCAAGAGAAAAAGGTAACTTCTCTGGAAGAACTTACTGAGTATATATTAAGTGGTG 2376  
Db 2216 ATATCAAGAGAAAAAGGTAACTTCTCTGGAAGAACTTACTGAGTATATTAAGTGGTG 2275  
Qy 2377 ATCAGGATCGCAAGGACCATGGGAAGAAATTTGCAAGGAGTGGATCGAGTTCCAGAAAC 2436  
Db 2276 ATCAGGATCGCAAGGACCATGGGAAGAAATTTGCAAGGAGTGGATCGAGTTCCAGAAAC 2335  
Qy 2437 GAATTCAGAAACAGAAAGTTGCCATTTGGCTACTGCTTGGCTTTGAGCAAGAAAGAAA 2496  
Db 2336 GAATTCAGAAACAGAAAGTTGCCATTTGGCTACTGCTTGGCTTTGAGCAAGAAAGAAA 2395  
Qy 2497 TGTTTAATAGACTGAGGATCGAGTCAAGGACTTACTGTTGTACAGCAATGTTTAAAGAGT 2556  
Db 2396 TGTTTAATAGACTGAGGATCGAGTCAAGGACTTACTGTTGTACAGCAATGTTTAAAGAGT 2455  
Qy 2557 GAAACAGCTGCAACCCGTGCCACTCTGCTCTTACTTGAAGATTTTCCCTTAAAAACAA 2616  
Db 2456 GAAACAGCTGCAACCCGTGCCACTCTGCTCTTACTTGAAGATTTTCCCTTAAAAACAA 2515  
Qy 2617 CACTGGCAGCTGTCTTGGACATGTTTTTAAAGAAACAACTTGTATCTAGAGATGCAAGTT 2676  
Db 2516 CACTGGCAGCTGTCTTGGACATGTTTTTAAAGAAACAACTTGTATCTAGAGATGCAAGTT 2575  
Qy 2677 TGATTAATTTTGGGTAAATGTGTCTCAATTAGAAACAACTCCGATAATGAAGAATCTCT 2736  
Db 2576 TGATTAATTTTGGGTAAATGTGTCTCAATTAGAAACAACTCCGATAATGAAGAATCTCT 2635

QY 2737 TATCTGTAATCCTCTCTTTTCCCTAATTTAGTTGGATGTGGGTTTGTCTTTTGTAGAGGGT 2796  
 Db |||||  
 2636 TATCTGTAATCCTCTCTTTTCCCTAATTTAGTTGGATGTGGGTTTGTCTTTTGTAGAGGGT 2695  
 QY |||||  
 2797 CTCACTGCATAAATCTGTGTTGGCCCTAGGCTATATGTATGTAGACCAAGGTTGGCTTTGAAAGT 2856  
 Db |||||  
 2696 CTCACTGCATAAATCTGTGTTGGCCCTAGGCTATATGTATGTAGACCAAGGTTGGCCCTTGAAGT 2755  
 QY |||||  
 2857 TGGTATTTCCCTTGGCCCTGTGTTTCCCTGTGCCACCAATGCCAGCTGGAAGTGTTTTAAATA 2916  
 Db |||||  
 2756 TGGTATTTCCCTTGGCCCTGTGTTTCCCTGTGCCACCAATGCCAGCTGGAAGTGTTTTAAATA 2815  
 QY |||||  
 2917 TCCTCTGCTTACCTGGGCTGAGAGTTTAAATTTTGCACTTTCAAAACGTTTCTCTGGAGG 2976  
 Db |||||  
 2816 TCCTCTGCTTACCTGGGCTGAGAGTTTAAATTTTGCACTTTCAAACGTTTCTCTGGAGG 2875  
 QY |||||  
 2977 CAGGGCTGGTCTGTTTACATACAGGCTTCAGGCCAGCTAGGGCTATGTGGTGAACCT 3036  
 Db |||||  
 2876 CAGGGCTGGTCTGTTTACATACAGGCTTCAGGCCAGCTAGGGCTATGTGGTGAACCT 2935  
 QY |||||  
 3037 AGTCTTAGAGGACAAACAGAACCAACAGTCAAGTCTAGGTTTACCTGGAACCTAGGCCAGGAGA 3096  
 Db |||||  
 2936 AGTCTTAGAGGACAAACAGAACCAACAGTCAAGTCTAGGTTTACCTGGAACCTAGGCCAGGAGA 2995  
 QY |||||  
 3097 TAGGAAGGTCAGGTATGCTTGGAACTCAAGTCACTCAAGGCCAGCATGGGGA 3156  
 Db |||||  
 2996 TAGGAAGGTCAGGTATGCTTGGAACTCAAGTCACTCAAGGCCAGCATGGGGA 3055  
 QY |||||  
 3157 GTTTAGCAAGACCTTTTACACTCAGAAATGGAAGAACCTGGGGATATAGCTCGGTAG 3216  
 Db |||||  
 3056 GTTTAGCAAGACCTTTTACACTCAGAAATGGAAGAACCTGGGGATATAGCTCGGTAG 3115  
 QY |||||  
 3217 CAGAGGCCCTGCCTACATGTGCAAGTCTCGGGTTCAAGTCCCGAGTACTCCAAATAGAA 3276  
 Db |||||  
 3116 CAGAGGCCCTGCCTACATGTGCAAGTCTCGGGTTCAAGTCCCGAGTACTCCAAATAGAA 3175  
 QY |||||  
 3277 AGAAAAACAATGTCTGGATAACTATAAGGTTTAAAGCTCATAGTCAAGTCTTAACCTCAA 3336  
 Db |||||  
 3176 AGAAAAACAATGTCTGGATAACTATAAGGTTTAAAGCTCATAGTCAAGTCTTAACCTCAA 3235  
 QY |||||  
 3337 ATTATGCAAGCAATGAGTCTTGTGCTGTTTTCTGTTTCTAAATTAAGGCTTTATGG 3396  
 Db |||||  
 3236 ATTATGCAAGCAATGAGTCTTGTGCTGTTTTCTGTTTCTAAATTAAGGCTTTATGG 3295  
 QY |||||  
 3397 GTTTTGTGTTTTGTATGTTTTATGCTGTATATTTTTCAGGTGAGGGATTCAGGCAGGGT 3456  
 Db |||||  
 3296 GTTTTGTGTTTTGTATGTTTTATGCTGTATATTTTTCAGGTGAGGGATTCAGGCAGGGT 3355  
 QY |||||  
 3457 CTCAGTGTGGCCTTGAGCTCATGGCAGTTTTTCTGCTTCAACTCAGTGTGTTTGTGTTT 3516  
 Db |||||  
 3356 CTCAGTGTGGCCTTGAGCTCATGGCAGTTTTTCTGCTTCAACTCAGTGTGTTTGTGTTT 3415  
 QY |||||  
 3517 TGTGTTTGTGTTTTTTTACTTTATAGATTTGACCTTAATGAAGGCAAAAACCTGTTATCA 3576  
 Db |||||  
 3416 TGTGTTTGTGTTTTTTTACTTTATAGATTTGACCTTAATGAAGGCAAAAACCTGTTATCA 3475  
 QY |||||  
 3577 ACCTAAAGAGACATGATGCTGTCACCTCAGTGTGATTTTCTCCCTCTTTTTTTTTTCC 3636  
 Db |||||  
 3476 ACCTAAAGAGACATGATGCTGTCACCTCAGTGTGATTTTCTCCCTCTTTTTTTTTTCC 3534  
 QY |||||  
 3637 CCCAGAGCTGAGGACCAAAACCCAGGGCTTTTGCACTTGTAGGCAAGCCCTCTACCACTGA 3696  
 Db |||||  
 3535 CCCAGAGCTGAGGACCAAAACCCAGGGCTTTTGCACTTGTAGGCAAGCCCTCTACCACTGA 3594  
 QY |||||  
 3697 GCTAAATCCCAAAACCAACCCCGCCCTTTTCTTTTTTTTAAAGACATGGTCTTATATAGT 3756  
 Db |||||  
 3595 GCTAAATCCCAAAACCAACCCCGCCCTTTTCTTTTTTTTAAAGACATGGTCTTATATAGT 3654  
 QY |||||  
 3757 CTAGGCTGGCTTAAACTCATTTACTTAGCAGGATGGCTTTTGAAGATCTCTCGTCTCT 3816  
 Db |||||  
 3655 CTAGGCTGGCTTAAACTCATTTACTTAGCAGGATGGCTTTTGAAGATCTCTCGTCTCT 3714  
 QY |||||  
 3817 CTTGTTGGCTTGCAATGGCATGTGCCATATACCCAGATTTTTTCTCTGTATCTAGGCTCT 3876

Db |||||  
 3715 CTTGTTTGGCTTGCAATGGCATGTGCCATATACCCAGATTTTTTCTCTGTATCTAGGTC- 3773  
 QY |||||  
 3877 TTTTAAAGATATTTTCCAAAGGACTGCCCAAAATCCACAGTGCAGTACTTTCTTGACTGGGA 3936  
 Db |||||  
 3774 TTTTAAAGATATTTTCCAAAGGACTGCCCAAAATCCACAGTGCAGTACTTTCTTGACTGGGA 3833  
 QY |||||  
 3937 AAGCGGGTGTGGTGTCTTTTCAAAGGCACACACCATATATGCCAGCTGGTCTGCCAAAC 3996  
 Db |||||  
 3834 AAGCGGGTGTGGTGTCTTTTCAAAGGCACACACCATATATGCCAGCTGGTCTGCCAAAC 3893  
 QY |||||  
 3997 TTAACATATTTGGTCTCTGTGCAAACTGTCCATTTGGAAGGTTTCTGTGTGTGTAGTTT 4056  
 Db |||||  
 3994 TTAACATATTTGGTCTCTGTGCAAACTGTCCATTTGGAAGGTTTCTGTGTGTGTAGTTT 3953  
 QY |||||  
 4057 CAGTTGAATGTGGCTCTCTGTAGCTTTCAAGAATGGGTAGANAATCATATAAGACACTCTTAA 4116  
 Db |||||  
 3954 CAGTTGAATGTGGCTCTCTGTAGCTTTCAAGAATGGGTAGANAATCATATAAGACACTCTTAA 4013  
 QY |||||  
 4117 GTATCATTTGATTTAGTGTCTTAAACACAGTGTCTTAAACACCCAGTGTACTTACACCCCTCAGCA 4176  
 Db |||||  
 4014 GTATCATTTGATTTAGTGTCTTAAACACAGTGTCTTAAACACCCAGTGTACTTACACCCCTCAGCA 4072  
 QY |||||  
 4177 AGATATTTCACTTTTGTGTTTATGTCTTAAACACAGTGTCTTAAACACCCCTCAGCA 4236  
 Db |||||  
 4073 AGATATTTCACTTTTGTGTTTATGTCTTAAACACAGTGTCTTAAACACCCCTCAGCA 4132  
 QY |||||  
 4237 CTGGCCTTTTCAAAACCTGTCTGGAGACACTGTAAACTTTGGATGGTGCAGTTCGCGTT 4296  
 Db |||||  
 4133 CTGGCCTTTTCAAAACCTGTCTGGAGACACTGTAAACTTTGGATGGTGCAGTTCGCGTT 4192  
 QY |||||  
 4297 TTTCAATGTTAAATGTCAACTCAAAAGGTCAATATCCAGGTTTGTGTGTCTTACCTTAT 4356  
 Db |||||  
 4193 TTTCAATGTTAAATGTCAACTCAAAAGGTCAATATCCAGGTTTGTGTGTCTTACCTTAT 4252  
 QY |||||  
 4357 GTAGAGAGAGCAAGCAAGCAAGGCGAGATAGAGCAGCCAGAAAGTCTAGTGT- CCCCC 4415  
 Db |||||  
 4253 GTAGAGAGAGCAAGCAAGCAAGGCGAGATAGAGCAGCCAGAAAGTCTAGTGTCCCCC 4312  
 QY |||||  
 4416 CCAAAGCGCTCTCTAGATAGTGTGGATCAGGTGTGCTTGGTTTGTTCAGTTAGGCTC 4475  
 Db |||||  
 4313 CCAAAGCGCTCTCTAGATAGTGTGGATCAGGTGTGCTTGGTTTGTTCAGTTAGGCTC 4372  
 QY |||||  
 4476 ACCTGTTGTGATGTTGAGTAAATGTGTTTACCAAGTCTTTTGTGTTTAAATGACAAAA 4535  
 Db |||||  
 4373 ACCTGTTGTGATGTTGAGTAAATGTGTTTACCAAGTCTTTTGTGTTTAAATGACAAAA 4432  
 QY |||||  
 4536 GAGAGGACCTCTGAGTGGGTTGGGATGAGCTTTCCAGACCTGGCAGGGTAAACTACCT 4595  
 Db |||||  
 4433 GAGAGGACCTCTGAGTGGGTTGGGATGAGCTTTCCAGACCTGGCAGGGTAAACTACCT 4492  
 QY |||||  
 4596 CAGTTTATATCTCCCTGTTTATTTCCGTTTGTGATCTTAAGTCTGCTCAGTGGTG 4655  
 Db |||||  
 4493 CAGTTTATATCTCCCTGTTTATTTCCGTTTGTGATCTTAAGTCTGCTCAGTGGTG 4552  
 QY |||||  
 4656 ATGATGTTTCAACACACCAAGGTTAGTAAGAGTGCACGACCCAGAAAAAGTGTGTTTAT 4715  
 Db |||||  
 4553 ATGATGTTTCAACACACCAAGGTTAGTAAGAGTGCACGACCCAGAAAAAGTGTGTTTAT 4612  
 QY |||||  
 4716 TTTGAGAACCCCAATTTCTGTGTTATTTTATGCACTTCCCTTTTGTGAACTCCAGAGTGCA 4775  
 Db |||||  
 4613 TTTGAGAACCCCAATTTCTGTGTTATTTTATGCACTTCCCTTTTGTGAACTCCAGAGTGCA 4672  
 QY |||||  
 4776 TTAAGAGTCTGTTTGTAGTCCCGTGGGAATGGGCTAGTTTGAAGCTATGTTTGGGAAGC 4835  
 Db |||||  
 4673 TTAAGAGTCTGTTTGTAGTCCCGTGGGAATGGGCTAGTTTGAAGCTATGTTTGGGAAGC 4732  
 QY |||||  
 4836 AGGCAAGTCTGACTTTAGGAAGAAAAAGTGTGACAGTGTGTAGACATTTCTTTTAAACCGG 4895  
 Db |||||  
 4733 AGGCAAGTCTGACTTTAGGAAGAAAAAGTGTGACAGTGTGTAGACATTTCTTTTAAACCGG 4792  
 QY |||||  
 4896 ACTGACGCTTAAACCACTTGTATTCAGATGTTAGGTTTGTGTTTCTGAGACCCAGGAC 4955





QY 2512 GGATCGAGTCAAAAGGACTTACTGTGTGTACAGCAATGTTAAGAAGTGAACAGCCTGCAACC 2571  
 DB 746 GGATCGAGTCAAAAGGACTTACTGTGTGTACAGCAATGTTAAGAAGTGAACAGCCTGCAACC 805  
 QY 2572 CGTGCCCACTCTGTCTCTTAATTGAGAGTTCCTTAAAAAACAAACACTGCGACCTGTCC 2631  
 DB 806 CGTGCCCACTCTGTCTCTTAATTGAGAGTTCCTTAAAAAACAAACACTGCGACCTGTCC 865  
 QY 2632 TTGGACATGTTTTTAAAGAAACAACTGTATCTAGAGATGACAGTTGATTTATTTTGGGT 2691  
 DB 866 TTGGACATGTTTTTAAAGAAACAACTGTATCTAGAGATGACAGTTGATTTATTTTGGGT 925  
 QY 2692 AATGTGTCTCATTTAGAAACAACTCCGATAAATGAAGAACTCTCTTATCTGTAAATCCTCT 2751  
 DB 926 AATGTGTCTCATTTAGAAACAACTCCGATAAATGAAGAACTCTCTTATCTGTAAATCCTCT 985  
 QY 2752 CTTTTTCTATTAGTTGGATGTGGTTTTGTCTTTTGGAGAGGTCTCACTGCAATAATTC 2811  
 DB 986 CTTTTTCTATTAGTTGGATGTGGTTTTGTCTTTTGGAGAGGTCTCACTGCAATAATTC 1045  
 QY 2812 TGTGTGGCCTAGGTCTATGTATGTAGACAGGTGGCCCTTGAAGTTGGTATTTCTTTGCC 2871  
 DB 1046 TGTGTGGCCTAGGTCTATGTATGTAGACAGGTGGCCCTTGAAGTTGGTATTTCTTTGCC 1105  
 QY 2872 TCTGTTTCTGTGCCACCATGCCAGCTGGAAAGTGTGTTTTAAATATCTCTGCTTACCTG 2931  
 DB 1106 TCTGTTTCTGTGCCACCATGCCAGCTGGAAAGTGTGTTTTAAATATCTCTGCTTACCTG 1165  
 QY 2932 GGGTGAGAGTTAAATTTTGTCACTTTTCAAAACGTTTTCTCTGGAGCAGGGGCTGGTCTGG 2991  
 DB 1166 GGGTGAGAGTTAAATTTTGTCACTTTTCAAAACGTTTTCTCTGGAGCAGGGGCTGGTCTGG 1225  
 QY 2992 TTTACATACAGGCTTCAGGCCAGCTAGGGCTATGTGTGAGACCTAGTCTTTAGAGGACAA 3051  
 DB 1226 TTTACATACAGGCTTCAGGCCAGCTAGGGCTATGTGTGAGACCTAGTCTTTAGAGGACAA 1285  
 QY 3052 ACAGAAACAAACAGTCAGGTTACTGTGGAACACTCAGGCAGGAGGATGAAGGTCCAGGT 3111  
 DB 1286 ACAGAAACAAACAGTCAGGTTACTGTGGAACACTCAGGCAGGAGGATGAAGGTCCAGGT 1345  
 QY 3112 ATGCTGGAACTCAGTGAACCTGAACTCAAGGCCAGCATGGGAGTTTACGAAGACCTT 3171  
 DB 1346 ATGCTGGAACTCAGTGAACCTGAACTCAAGGCCAGCATGGGAGTTTACGAAGACCTT 1405  
 QY 3172 TTTACATCAGAAATGGAAAGGAGCTGGGATATAGCTCGGTAGCAGAGGCCCTGCCT 3231  
 DB 1406 TTTACATCAGAAATGGAAAGGAGCTGGGATATAGCTCGGTAGCAGAGGCCCTGCCT 1465  
 QY 3232 ACATGTGCAAGTCTCTGGTTTCACTCCCGAGTACTGCAAAATAGAAAGAAAAACATTTGTC 3291  
 DB 1466 ACATGTGCAAGTCTCTGGTTTCACTCCCGAGTACTGCAAAATAGAAAGAAAAACATTTGTC 1525  
 QY 3292 TTGGATAACTATAAGGTTTAAAGCCTCATAGTCACTTCAAAATATGCAATGCAATTG 3351  
 DB 1526 TTGGATAACTATAAGGTTTAAAGCCTCATAGTCACTTCAAAATATGCAATGCAATTG 1585  
 QY 3352 AGTTCTGTGTGCTTTTCTGTTCTTAAATTAATGGGCTTTATGGGTTTGTGTTTGTAT 3411  
 DB 1586 AGTTCTGTGTGCTTTTCTGTTCTTAAATTAATGGGCTTTATGGGTTTGTGTTTGTAT 1645  
 QY 3412 GTTTTTATGTCTGTTATATTTTGAAGTAGGGATTTAGGCCAGGGTCTCACTGTGGCCCTTG 3471  
 DB 1646 GTTTTTATGTCTGTTATATTTTGAAGTAGGGATTTAGGCCAGGGTCTCACTGTGGCCCTTG 1705  
 QY 3472 AGCTCATGGCAGTTTTCTGCTTCAACTGAGTGTGTTTTGTTTTGTTTTGTTTTT 3531  
 DB 1706 AGCTCATGGCAGTTTTCTGCTTCAACTGAGTGTGTTTTGTTTTGTTTTGTTTTT 1765  
 QY 3532 TTACTTTTCATAGATTTTGAATTAAGAGGCAAAACCTGTTATCAACCTAAAGACATTG 3591  
 DB 1766 TTACTTTTCATAGATTTTGAATTAAGAGGCAAAACCTGTTATCAACCTAAAGACATTG 1825

QY 3592 ATGTGTCACTTCAGTGGTGGATTTTCTCCCTC-TTTTTTTTTTTTTTCCCCCAGAGCTGAGGA 3650  
 DB 1826 ATGTGTCACTTCAGTGGTGGATTTTCTCCCTC-TTTTTTTTTTTTTTCCCCCAGAGCTGAGGA 1885  
 QY 3651 CAAAACCCAGGGCTTTTGCACCTGTAGGCAAGGCTCTACCACTGAGCTAAATCCCAAA 3710  
 DB 1886 CAAAACCCAGGGCTTTTGCACCTGTAGGCAAGGCTCTACCACTGAGCTAAATCCCAAA 1945  
 QY 3711 CCA-CCCCCCCCCTTTTCTTTTTTAAAGACATGGTCTTATATAGTCTTAGGCTGGGCTT 3769  
 DB 1946 CCACCCCCCCCCCTTTTCTTTTTTAAAGACATGGTCTTATATAGTCTTAGGCTGGGCTT 2005  
 QY 3770 AAACTCAATTAATTAGCCAAAGGATGGCTTTGAAAGATCCTCTGTCTCTTTGTTGGCTTG 3829  
 DB 2006 AAACTCAATTAATTAGCCAAAGGATGGCTTTGAAAGATCCTCTGTCTCTTTGTTGGCTTG 2065  
 QY 3830 CACATGCGCATGTGCGCATACATCCAGATTTTTCTCTGTATCTAGGTCTTTTTTAAGATATTT 3889  
 DB 2066 CACATGCGCATGTGCGCATACATCCAGATTTTTCTCTGTATCTAGGTCTTTTTTAAGATATTT 2125  
 QY 3890 CCAAGGACTGCCACAAATCCACAGTCAGTACTTTCTTGACCTGGGAAAGCGGGTTGGT 3949  
 DB 2126 CCAAGGACTGCCACAAATCCACAGTCAGTACTTTCTTGACCTGGGAAAGCGGGTTGGT 2185  
 QY 3950 GTGCTTTTCAAAGGCACACACCAATTAATGCCCAGCTGGTCTGCCAAACTTAACATATTTGG 4009  
 DB 2186 GTGCTTTTCAAAGGCACACACCAATTAATGCCCAGCTGGTCTGCCAAACTTAACATATTTGG 2245  
 QY 4010 TTTCTGTGCAAACTGTCCATTTTGAAGGTTTCTGTGTGTTGTAGTTTTCAGTTGAAATGG 4069  
 DB 2246 TTTCTGTGCAAACTGTCCATTTTGAAGGTTTCTGTGTGTTGTAGTTTTCAGTTGAAATGG 2305  
 QY 4070 CTCTCTGTAGCTTTCAAGAATGGTAGAAATCATAAAGCACTCTTAAGTAAATCATTTGCAT 4129  
 DB 2306 CTCTCTGTAGCTTTCAAGAATGGTAGAAATCATAAAGCACTCTTAAGTAAATCATTTGCAT 2365  
 QY 4130 TGTAGACATTTTTTTTTTAACTAGGGGTAATTTGGGAGACATGTCCAGATATTTCACTTT 4189  
 DB 2366 TGTAGACATTTTTTTTTTAACTAGGGGTAATTTGGGAGACATGTCCAGATATTTCACTTT 2425  
 QY 4190 TGTATTATGTCTTAAACACAGTTACTTACACCTGCTCAGCACTGGCAGCTTTTCA 4249  
 DB 2426 TGTATTATGTCTTAAACACAGTTACTTACACCTGCTCAGCACTGGCAGCTTTTCA 2485  
 QY 4250 AACCTGTCTTGGAGACACTGTAACTTGGATGGTGCAGTTCTGGGTTTTCATGTGTAAAA 4309  
 DB 2486 AACCTGTCTTGGAGACACTGTAACTTGGATGGTGCAGTTCTGGGTTTTCATGTGTAAAA 2545  
 QY 4310 TGTCAACTACAAAGGTCAATATATCCAGGTTTGTGTCTTACCTTATGTAGAGAGAGCA 4369  
 DB 2546 TGTCAACTACAAAGGTCAATATATCCAGGTTTGTGTCTTACCTTATGTAGAGAGAGCA 2605  
 QY 4370 AAGCAGAAAGGCGCAGATAGAGCGAGCAGAAAGTGTAGTGTCCCCCCCCAAAGCGTGTTC 4429  
 DB 2606 AAGCAGAAAGGCGCAGATAGAGCGAGCAGAAAGTGTAGTGTCCCCCCCCAAAGCGTGTTC 2665  
 QY 4430 TAGATGTGTGTGATCAGGTGTCTGTGTTTGTTCAGTTAGCGTCACTGTGTGTGCATG 4489  
 DB 2666 TAGATGTGTGTGATCAGGTGTCTGTGTTTGTTCAGTTAGCGTCACTGTGTGTGCATG 2725  
 QY 4490 GTTGAGGTAAATGTGTTTCAAGTTCTTTAGTGGTTTACATGCAAAAAGAGAGGACCTCTGA 4549  
 DB 2726 GTTGAGGTAAATGTGTTTCAAGTTCTTTAGTGGTTTACATGCAAAAAGAGAGGACCTCTGA 2785  
 QY 4550 GTGGGTGTGGGATGAGCTTTCCAGACCTGGAGGGTAAACTACCTCAGTTTATAATCTC 4609  
 DB 2786 GTGGGTGTGGGATGAGCTTTCCAGACCTGGAGGGTAAACTACCTCAGTTTATAATCTC 2845  
 QY 4610 CCTGGTTATTTTCCGTTTGTATGTGATCTTAAGGTCTGCGCTCAGTGGTGTATGTTTCATCCA 4669  
 DB 2846 CCTGGTTATTTTCCGTTTGTATGTGATCTTAAGGTCTGCGCTCAGTGGTGTATGTTTCATCCA 2905  
 QY 4670 CACACAGAGGTTAGTAAGAGTGCACGACCAAGAAACGTTGGTCTTTATTTTTTGGAGAACCCCA 4729



```
Db 2906 CACACAAAGGTTAGTAAGAGTGTCACGACGAGAAACGTTGGTCTTATTTTGTGAGACCCCA 2965
Qy 4730 TTTCTGTGTTATTTATGCACTGCTTCTAGTGAACCTCAGAGTGCATTAAGAGTCTGGT 4789
Db 2966 TTTCTGTGTTATTTATGCACTGCTTCTAGTGAACCTCAGAGTGCATTAAGAGTCTGGT 3025
Qy 4790 TTAGTGCCGTGGGAATGGCTAGTTTGAAGCTATGTTTGGAAAGCAGGCAAGTTGACTT 4849
Db 3026 TTAGTGCCGTGGGAATGGCTAGTTTGAAGCTATGTTTGGAAAGCAGGCAAGTTGACTT 3085
Qy 4850 TAGGAAGAAAGCTGTGACAGTGTGTAGACATTTCTTTTAAACCGGACTGAGCTTAACA 4909
Db 3086 TAGGAAGAAAGCTGTGACAGTGTGTAGACATTTCTTTTAAACCGGACTGAGCTTAACA 3145
Qy 4910 ACACCTGATTTTCAGATGATAGGTTTGTGTTCTGAGACCCAGCACCTGTATATTTAAAA 4969
Db 3146 ACACCTGATTTTCAGATGATAGGTTTGTGTTCTGAGACCCAGCACCTGTATATTTAAAA 3205
Qy 4970 ATTGTTCCAGATTACACCTTCACATCAATAGTAATGACTCATGCTGCGACACATGT 5029
Db 3206 ATTGTTCCAGATTACACCTTCACATCAATAGTAATGACTCATGCTGCGACACATGT 3265
Qy 5030 CCTGATGTTGTCAGACAGAGGATCTTTGACTGAGGAGAGAAAACTGTCTATTTGTCATC 5089
Db 3266 CCTGATGTTGTCAGACAGAGGATCTTTGACTGAGGAGAGAAAACTGTCTATTTGTCATC 3325
Qy 5090 CCAGCCCCCAGGAAGAACCTCCAGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 5149
Db 3326 CCAGCCCCCAGGAAGAACCTCCAGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3385
Qy 5150 TAGTTGAATACACATTCAGTCTTGCAGTGGTCTTTAGATCTGTGTAGCATGTGAGGCT 5209
Db 3386 TAGTTGAATACACATTCAGTCTTGCAGTGGTCTTTAGATCTGTGTAGCATGTGAGGCT 3445
Qy 5210 CTGTACAGTGGGGGCGACACTCTGAGGCTGGAATGTGCGACACCTTTATCTAACTTG 5269
Db 3446 CTGTACAGTGGGGGCGACACTCTGAGGCTGGAATGTGCGACACCTTTATCTAACTTG 3505
Qy 5270 AAATCAAAACCGTCACAAATTTTATTTTATTAATTTAAGAAAGAGTTGGGAATGACATTT 5329
Db 3506 AAATCAAAACCGTCACAAATTTTATTTTATTAATTTAAGAAAGAGTTGGGAATGACATTT 3565
Qy 5330 TTTGAGTTGGCTTTTCAGCTCAGTCATTTTACGTGTAACGTGGAGATTTGATAGCTCAG 5389
Db 3566 TTTGAGTTGGCTTTTCAGCTCAGTCATTTTACGTGTAACGTGGAGATTTGATAGCTCAG 3625
Qy 5390 ATTATATTGTATATATTAATTAATCTGTAATCTGTAATGTAATTAATTAATTTGCAATTA 5449
Db 3626 ATTATATTGTATATATTAATTAATCTGTAATCTGTAATGTAATTAATTAATTTGCAATTA 3685
Qy 5450 TTAATAAAAAAAAAAAAAA 5468
Db 3686 TTAATAAAAAAAAAAAAAA 3704

RESULT 3
AC107842 215781 bp DNA linear ROD 22-OCT-2003
LOCUS Mus musculus chromosome 1, clone RP23-30409, complete sequence.
DEFINITION AC107842
ACCESSION AC107842.13 GI:37806519
VERSION AC107842.13
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 215781)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS Birren, B., Nusbaum, C. and Lander, E.
TITLES Mus musculus chromosome 1, clone RP23-30409
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 215781)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
```

```
Anderson, S., Barna, N., Bastien, V., Boguslavskiy, L., Boukhgalter, B.,
Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,
Choepe, Y., Collangelo, M., Collins, S., Collymore, A., Cook, A.,
Cooke, P., DeArellano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S.,
Farrisa, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S.,
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Jones,
Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
Kamat, A., Karatas, A., Kells, C., LaRoque, K., Lamazares, R.,
Landers, T., Lechack, J., Levine, R., Liu, G., MacLean, C.,
Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M.,
McEwan, P., McKernan, K., Meldrim, J., Meneus, L., Mihova, T.,
Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C.,
Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J.,
Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C.,
Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J.,
Rosetti, M., Roy, A., Santos, R., Schauer, S., Schuback, R., Seaman, S.,
Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Trigglio, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (24-JAN-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 215781)
Anderson, M., Arachchi, H. M., Barna, N., Bastien, V., Bloom, T.,
Boguslavskiy, L., Boukhgalter, B., Camarata, J., Chang, J., Choepe, Y.,
Collymore, A., Cook, A., Cooke, P., Corum, B., DeArellano, K.,
Diaz, J. S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S.,
Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S.,
Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B.,
Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R.,
Lindblad-Toh, K., Liu, G., Lui, X., Mabbitt, R., MacLean, C.,
Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M.,
Meldrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J.,
Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P.,
O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N.,
Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P.,
Roman, J., Schauer, S., Schuback, R., Seaman, S., Severy, P., Smith, C.,
Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M.,
Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M.,
Vassiliev, H., Venkataraman, V. S., Viel, R., Vo, A., Wilson, B., Wu, X.,
Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (21-OCT-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
4 (bases 1 to 215781)
Anderson, M., Arachchi, H. M., Barna, N., Bastien, V., Bloom, T.,
Boguslavskiy, L., Boukhgalter, B., Camarata, J., Chang, J., Choepe, Y.,
Collymore, A., Cook, A., Cooke, P., Corum, B., DeArellano, K.,
Diaz, J. S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S.,
Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S.,
Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B.,
Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R.,
Lindblad-Toh, K., Liu, G., Lui, X., Mabbitt, R., MacLean, C.,
Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M.,
Meldrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J.,
Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P.,
O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N.,
Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P.,
Roman, J., Schauer, S., Schuback, R., Seaman, S., Severy, P., Smith, C.,
Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M.,
Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M.,
Vassiliev, H., Venkataraman, V. S., Viel, R., Vo, A., Wilson, B., Wu, X.,
Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (22-OCT-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Oct 22, 2003 this sequence version replaced gi:3777388.
```

All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

----- Project Information

Center project name: L20605

Center clone name: 304\_O\_9

-----

# FEATURES

## source

Location/Qualifiers

1..215781  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/chromosome="1"  
/map="1"  
/clone="RP23-304O9"  
/clone.lib="RPCI-23 Female Mouse BAC"

## misc\_feature

1..8553  
/note="wgs\_end\_extension"

clone\_end:SP6"

## repeat\_region

complement(2..143)

## repeat\_region

/rpt\_family="URRIA"

## repeat\_region

441..478

## repeat\_region

/rpt\_family="(TTG)n"

## repeat\_region

776..815

## repeat\_region

/rpt\_family="(CA)n"

## repeat\_region

complement(817..1006)

## repeat\_region

/rpt\_family="ID\_B1"

## repeat\_region

complement(1092..1181)

## repeat\_region

/rpt\_family="ID3"

## repeat\_region

1304..1512

## repeat\_region

/rpt\_family="B3"

## repeat\_region

complement(1754..1928)

## repeat\_region

/rpt\_family="B3"

## repeat\_region

complement(2578..2681)

## repeat\_region

/rpt\_family="MIR"

## repeat\_region

2809..2907

## repeat\_region

/rpt\_family="B1\_MM"

## repeat\_region

2902..2957

## repeat\_region

/rpt\_family="B2\_Mm1"

## repeat\_region

complement(2986..3154)

## repeat\_region

/rpt\_family="B3A"

## repeat\_region

complement(3608..3753)

## repeat\_region

/rpt\_family="B1\_MM"

## repeat\_region

complement(3772..3920)

## repeat\_region

/rpt\_family="B3A"

## repeat\_region

complement(3974..4059)

## repeat\_region

/rpt\_family="PB1D10"

## repeat\_region

complement(4101..4207)

## repeat\_region

/rpt\_family="PB1D9"

## repeat\_region

4828..4971

## repeat\_region

/rpt\_family="B1F"

## repeat\_region

5328..5474

## repeat\_region

/rpt\_family="B1\_MM"

## repeat\_region

complement(6606..6614)

## repeat\_region

/note="<30 qual SNGL region"

## repeat\_region

complement(6658..6664)

## repeat\_region

/note="<30 qual SNGL region"

## repeat\_region

complement(6700..6708)

## repeat\_region

/note="<30 qual SNGL region"

## repeat\_region

complement(6737..6743)

## repeat\_region

/note="<30 qual SNGL region"

## repeat\_region

complement(6823..6844)

## repeat\_region

/note="<30 qual SNGL region"

## repeat\_region

complement(6896..6900)

## repeat\_region

/note="<30 qual SNGL region"

## repeat\_region

7252..7338

## repeat\_region

/rpt\_family="B1F"

## repeat\_region

8546..8671

misc\_feature  
/rpt\_family="B1F"  
8554..8559  
/note="clone\_boundary"  
clone\_end:SP6  
site:Boori"  
9091..9403  
/rpt\_family="ORR1D"  
9406..9777  
/rpt\_family="ORR1B-int"  
9866..10062  
/rpt\_family="ORR1D"  
10732..10821  
/rpt\_family="ID\_B1"  
complement(11142..11221)  
/rpt\_family="PB1D9"  
12068..12243  
/rpt\_family="B2\_Mm2"  
12341..12406  
/rpt\_family="ORR1D"  
12413..12778  
/rpt\_family="ORR1C"  
complement(13030..13132)  
/rpt\_family="B1\_MM"  
complement(13345..13408)  
/rpt\_family="RSINE1"  
complement(13429..13690)  
/rpt\_family="B4"  
13830..13980  
/rpt\_family="B1\_MM"  
14513..14624  
/rpt\_family="PB1D7"  
14642..14699  
/rpt\_family="G-rich"  
14860..14963  
/rpt\_family="PB1D9"  
complement(15187..15305)  
/rpt\_family="B3"  
complement(15492..15644)  
/rpt\_family="B1\_MM"

Query Match 56.7%; Score 3100; DB 10; Length 215781;

Best Local Similarity 99.9%; Pred. No. 0;

Matches 3122; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

Qy 2332 AGGTAACTCTCTGGAGAAGTTACTGAGTATTATATAAGTGTGATGAGATCGCAAAG 2391  
Db 156280 AGGTAACTCTCTGGAGAAGTTACTGAGTATTATATAAGTGTGATGAGATCGCAAAG 156339  
Qy 2392 GACCATGGGAAGAAATTTGCAAGGGATGGATGCGAGTTCCAGAAACGAATTCAGAAACAG 2451  
Db 156340 GACCATGGGAAGAAATTTGCAAGGGATGGATGCGAGTTCCAGAAACGAATTCAGAAACAG 156399  
Qy 2452 AAGTTGCCATTGGCTACTGCTGGCCCTTTGAGCAGACAGAGAAAATGTTTAAATAGACTGA 2511  
Db 156400 AAGTTGCCATTGGCTACTGCTGGCCCTTTGAGCAGACAGAGAAAATGTTTAAATAGACTGA 156459  
Qy 2512 GGATCGAGTCAAAGGACTTACTGTTGACAGCAATGTTAAGAAGTGAACAGCCCTGCAACC 2571  
Db 156460 GGATCGAGTCAAAGGACTTACTGTTGACAGCAATGTTAAGAAGTGAACAGCCCTGCAACC 156519  
Qy 2572 CGTGCCCACTCTGCTCTTACTTTGAGAGTTTCCCTTAAAAACAAACACCTGGCAGCTGTCC 2631  
Db 156520 CGTGCCCACTCTGCTCTTACTTTGAGAGTTTCCCTTAAAAACAAACACCTGGCAGCTGTCC 156579  
Qy 2632 TTGGACATGTTTTTAAAGAACCAACTGTATCTTAGAGTCAGCTTTGATTTTGGGT 2691  
Db 156580 TTGGACATGTTTTTAAAGAACCAACTGTATCTTAGAGTCAGCTTTGATTTTGGGT 156639  
Qy 2692 AATGTGTCTCATTTAGAAACACCAACTCCGATAATGAAGATCTCTTATCTGTAACTCTCT 2751  
Db 156640 AATGTGTCTCATTTAGAAACACCAACTCCGATAATGAAGATCTCTTATCTGTAACTCTCT 156699  
Qy 2752 CTTTTCTATTTAGTTGGATGTGGGTTTTGTCTTTTTTGAGGGGTCTCACTGCATATTC 2811



QY	4971	TTGTTCCAGATTACACCTTCTCACTATCAAAATGAGTAAATGA	CTCATGCGCTGCAGACATGTC	5030
Db	158919	TTGTTCCAGATTACACCTTCTCACTATCAAAATGAGTAAATGA	CTCATGCGCTGCAGACATGTC	158978
QY	5031	CTGATGTTGGCAAGAACAGAGGATCTTTGACTGAAGGAGAA	AAACTCTGCTATGTCATCC	5090
Db	158979	CTGATGTTGGCAAGAACAGAGGATCTTTGACTGAAGGAGAA	AAACTCTGCTATGTCATCC	159038
QY	5091	CAGCCCCCAGGAAAGAACACCTTCCAGGCGAGGCGAGG	CAGGCGATGGTGGTCT	5150
Db	159039	CAGCCCCCAGGAAAGAACACCTTCCAGGCGAGGCGAGG	CAGGCGATGGTGGTCT	159098
QY	5151	AGTTGAATACACATTTCAAGCTTTGCAAGTGGTCTTTAG	ATCTGTAGCATGTGAGGCTC	5210
Db	159099	AGTTGAATACACATTTCAAGCTTTGCAAGTGGTCTTTAG	ATCTGTAGCATGTGAGGCTC	159158
QY	5211	TGTACAGGTGGGGGCCACACCTTCGAGGGCTGAAATATG	GCGAACCTTTATCTAACTTGA	5270
Db	159159	TGTACAGGTGGGGGCCACACCTTCGAGGGCTGAAATATG	GCGAACCTTTATCTAACTTGA	159218
QY	5271	AATCAAAACCGTCAAAATTTTTATTTTTATTAATTTTA	AGAAAGAGTTGGGGAATGACATTTT	5330
Db	159219	AATCAAAACCGTCAAAATTTTTATTTTTATTTTAAATTT	TAAGAAAGAGTTGGGGAATGACATTTT	159278
QY	5331	TTGAGTTGGCCCTTTTCAGCTCAGTCAATTTTACGTGTA	ACGTGGAGATTTGATAGCTCAGA	5390
Db	159279	TTGAGTTGGCCCTTTTCAGCTCAGTCAATTTTACGTGTA	ACGTGGAGATTTGATAGCTCAGA	159338
QY	5391	TTATATTGTATATAATTTATTAACATACTGTAATCTGT	ATAATTAATATATTGCAATTTAT	5450
Db	159339	TTATATTGTATATAATTTATTAACATACTGTAATCTGT	ATAATTAATATATTGCAATTTAT	159398
QY	5451	TAAA	5454	
Db	159399	TAAA	159402	
RESULT 4				
AC119774				
LOCUS				
DEFINITION	AC119774	212261 bp	DNA	linear
				HTG 19-NOV-2002
ACCESSION				
VERSION	AC119774.6	GI:25095262		
KEYWORDS	HTG; HTGS PHASE3; HTGS DRAFT; HTGS_FULLTOP.			
SOURCE	Rattus norvegicus (Norway rat)			
ORGANISM	Rattus norvegicus			
	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;			
	Rattus.			
REFERENCE				
AUTHORS	1 (bases 1 to 212261)			
	Muzny,D,Marie., Metzker,M,Lee., Abramzon,S., Adams,C., Alder,J.,			
	Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,			
	Anylalabechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,			
	Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,			
	Blawie,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,			
	Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,			
	Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,			
	Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,			
	Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,			
	Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,			
	Delgado,O., Denison,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,			
	Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,			
	Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,			
	Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,			
	Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,			
	Georgiev,G., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,			
	Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,			
	Harvey,Y., Harkins,P., Hawes,A., Henderson,N., Hernandez,J.,			
	Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,			
	Hollins,B., Howells,S., Hulyk,S., Hume,J., Idelbird,D., Jackson,A.,			
	Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,			

\* NOTE: This is a 'working draft' sequence. It currently consists of 2 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

\* 1 11068: contig of 11068 bp in length  
 \* 11069 11168: gap of unknown length  
 \* 11169 212261: contig of 201093 bp in length.

FEATURES

source Location/Qualifiers  
 1..212261  
 /organism="Rattus norvegicus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10116"  
 /clone="CH230-146P11"  
 misc\_feature 1..1217  
 /note="wgs\_contig"

ORIGIN

Query Match 44.1%; Score 2412.4; DB 2; Length 212261;  
 Best Local Similarity 75.7%; Pred. No. 0;  
 Matches 3968; Conservative 0; Mismatches 771; Indels 503; Gaps 59;  
 Qy 404 GCGTGACGCGCTGAGCTCTGTCTCTCTCTGTCTGAGAACGCCCAAGGAAAGGAGAT 463  
 Db 140602 GCTCTGATGCGCTGAGCTC--CGTACTCTGCGCGTGAAGTTGCCAAGGAAAGGAGAT 140658  
 Qy 464 GGAGACAGAACGACAGGCGCGGAGCGGCTGCGCTGCGCTGGCTGGCTCTGGTTCG 523  
 Db 140659 GGAGACGAGAACACACAGGGCTC-AAAGCGCGCTACCTCTCGATGGGCTCTGATTATG 140717  
 Qy 524 GCTGCGCTTCTCTGCGGATCGACGCTGCTCTCGAGTTCCGCGCGCTCTCTCG 583  
 Db 140718 CTGCGCTTCTTAGGGCGATCACGCTGCGACTCTGAGTTCCGAAACCTTCTTCTCA 140777  
 Qy 584 ACATAATCCCGGAACTCCGCTCTGCGGAGCGTGGACCAAGTACTGGACCAATTCCT 643  
 Db 140778 ACCAAATCTCGAGACTCTGCGCTGGTCCGAGCCTCAGACCAAGTACTGGACCAATTCGA 140837  
 Qy 644 TTCTCAGCTCTTGGCTGCTCTCTAGCTATTCAGAGTCTGCTTTGGAGCGAGCT 703  
 Db 140838 TTCCCACTTCTTGC---TCTCCCTAGCTTCTCTCAGAGTCTGCTATTTGGAGTCACT 140894  
 Qy 704 TTCCGGGGCTGATTCCTACAGATGCTAGATTTTCCGCAAGTTACAGCGCCCTGAG 763  
 Db 140895 TTTGGTGGCAATGATTCTGACCAAGATGCTATCTTTACAGAAATTACAGTCCCTGCG 140954  
 Qy 764 AGCTTCGAGAGGACGGGAGG--AATCTGACGCTCCCAAGTCTCTGAGTTACA 822  
 Db 140955 AGCTTTGAGAGGACGGGAGGAAATCAGCGCTCCCAAGTCTCTGAGTT-CT 141013  
 Qy 823 CTGCGCTGAGACTCTTCCGGAAGACTCGCGTGTGAGTACTCTTGCATGGCTAGAGAG 882  
 Db 141014 CTGCGCTGGAATTTTCTTCGGAAGACTTGGCTGTGAGTCTCTTGTATCGGCTAGAGAC 141073  
 Qy 883 GGACTCCAGTGGAGTCTGCTCTCTCAGCTGGAAGTTAAACTCAAGCCCGAGGAAAGAG 942  
 Db 141074 GGACTCCAGTGGAGTCTGCTCTCTCAGCTGGAATGAACTCAAGCCCGAGGAAAGAG 141133  
 Qy 943 CTTTAGACTCTGCAGCGCCCACTTTCTCTCTGAGCAGCAGCTGTGGGGAGTGGAGTTGC 1002  
 Db 141134 CTTTAGACTCTACAGCGCACACTTTCTCTCTATGGAGCAGCAGCTGTGGGGAGTGGAGTTGC 141193  
 Qy 1003 TGCCAGTAGGCTTCAAGCTGCTAGTCTCCCAACCGAGAACTTGACTCTTCATCTCTG 1062  
 Db 141194 TGCCCGCTCTCTTCAAAACCGATCTAGTCTCCCAACCGAGAACTTGGCTCTTCATCTCTG 141253  
 Qy 1063 GGCTCTGAGGCTCAGAGCTTAGTAATTTCAAGGTAGTTTCTATCTCTGAGCAACCTTT 1122  
 Db 141254 GGCTCTGAGCACTCAGAGCTTGGGCAATTTCAAGGTAGTTTCTATCTCTGAGCAACCTTT 141313

Qy 1123 CTTACTTGAGACTACCTTTCCCGCTAGGAGTGGCTGTCTGAGAGAGCGCTGAGAGTGGCC 1182  
 Db 141314 CTTACTTGAGACTACCTTTCCCGCTAGGAGTGGCTGTCTGAGAGAGCGCTGAGAGTGGCC 141373  
 Qy 1183 AGTTTGTGGCTTTCCGAACTAACCCACAGAGCTGCTATCTTCTGAAAGATGTTGTGTC 1242  
 Db 141374 AACTTGTGGATTTCCGAGCGCTATCCACAGAGCTGTTTACCTTTCTGAAAGTGGTGTGTC 141433  
 Qy 1243 ACCCTCAGCGCTTTCCGCGCAGAGATGTGCGCAACCGCTCGGAGAAGGTGTCCGCTCTCT 1302  
 Db 141434 ACTC-CAGCGCTGCGTGTGAGATGTGCGCAACTGCTGCGGAGGTGTCCAGCCCTCT 141492  
 Qy 1303 CTACAGAGCGCTGCGGAAATCCACACCTCTGATGAAACCGCTAGAAATTCCTCAGG 1362  
 Db 141493 CTACAGAGCGCTGCGGAAATCCACACCTCTGATGAAACCGCTAGAAATTCCTCAG- 141551  
 Qy 1363 CTAACAAAGGCAAGAGTTACCCACCCCTGACCAAGATATGCTATCATAGCTCGAGG 1422  
 Db 141552 -----AGGCAAGAGTTTACCAACCCCTGACCAAGCAATGGCTATCATAGCTCGAGG 141603  
 Qy 1423 AGGAACAATACTTCTCCGATGGAGCCCAACAATTCACAGATAACCCACAGCAGCGG 1482  
 Db 141604 AGGAGCACTGCTTCTCCGATGGAGCCCGCAACATTCACAGATAAGGTAGCAGGCTA 141663  
 Qy 1483 TGTCCCTCTGTGACAGAGCGCGGAGCCCACTGAGAAAAAACAGAAATTTGGTATTCAG 1542  
 Db 141664 TGTCCCTCTGTGAGCGAGCGCGGAGCCCACTGAGAAAAAAGTAGAAATTTGGTGTGAGG 141723  
 Qy 1543 AAGTTTACAGAGCGCCCAAGGAGCAGTCTGTTTGTGAATTTACCGTGGAAAAAGAAAT 1602  
 Db 141724 AAGTTTACAGAGCGCCCAAGGAGCAGTCTGCTTGTGAGTTAAGTGTGCAAAAAAGAAAT 141783  
 Qy 1603 GTGAAGAGGACCACTAATGCAACTGACCTCTCAGATAGAGAGAGAGCTTCTCTCTTT 1662  
 Db 141784 GTGAAGAGGACCACTAATGTAATTTGACTACTCAGACATAGGGGAAACCTTCTCTTT 141843  
 Qy 1663 CTACACAGACAGTTTGTAGCAAACTGATAGATTAATTTTGGAGGCGCCCGCAGTG 1722  
 Db 141844 CTCCAGACCGCTTGTACCAAACTGATAGATTAATTTTGGGAGGAGCGCTCCAGTG 141903  
 Qy 1723 ACTTGAAGCGCTGATTTCTGAAGTGAAGTGGGCGAGGACCTGAGAGCGATG 1782  
 Db 141904 ACTTGAAGCGCTGATTTCTGAAGTGAAGTGGGATGAGGAACTTGAATGATGATG 141963  
 Qy 1783 GCTTTGATAGCGATGGCTCCCTGTCTGAATCAGACGCTGGAACAGGACTCCGGAAGCGCTTC 1842  
 Db 141964 GTTTGATAGTGGCTCCCTGTCTGAATCAGACGGAACAGACTCCGGAAGCGCTTC 142023  
 Qy 1843 ACCTTTGGAACCTTTTCCACAGTGTAGATCCTTACAAACCCCAAACTTTTACAGCCAGA 1902  
 Db 142024 ACCTTTGGAACCTTTTCTACAGTGTAGATCCTTACAAACCCCAACACTTTTACAGCCAGA 142083  
 Qy 1903 TTCAAGCGCTGCGAGAAATTTGCGGAGAGCGCTTCCGAGAGACCCCGAGCCATAGTT 1962  
 Db 142084 TTCAAGCTGTGCGAGAAATTTGCGGAGAGCGCTTCCGAGAGACCCCGAGCCATAGTT 142143  
 Qy 1963 GCAGCTGTGTGTAGGAGCTGTGAGGAGGAGCGCTTCCGAGAGACCCCGAGCCATAGTT 2022  
 Db 142144 GCAACCGGATGTAGGAGGCTCTGAGGCACTCTCTTCCGAGAGCCCGTGAAGAGTT 142203  
 Qy 2023 CCGGGAGGAGAGTACTGGAAACCGAGTGCAGATGAAGCAGAGAAATCTTAAATTTGTGGA 2082  
 Db 142204 CTGGGAGGAGGAGTGAAGTGGAAACCTAGTGCAGAGGAGCAGAAAAATCATAGATTTGTGGA 142263  
 Qy 2083 ACTCTTCTGTCATTTCTGAGGAGCGCTTACAACTTTTAAATTTTAAAGGCTCTTTTCAAC 2142  
 Db 142264 ACTCTCTCTGTAATTTCTGAGGAGCGCTTACAACTTTTAAATTTTAAAGGCTCTTTCCAA 142323  
 Qy 2143 CGTCAGGAGAAATTTGAAAGCGCTCAGGACTCAAAAGGCTCTTCTGAGGCTCAGAGTGG 2202  
 Db 142324 CGTTAGGAAAGAGCTGGAAGGCTGTGAGGACTCTGGAAGGACTCTCTGGGGCCACAGTGG 142383  
 Qy 2203 CTTTCTTGCGCCATCATACCTTACTTTTGTAAAGGCCA-----GCTGTTAGAGAGCC 2256





QY 4347 TCTACCTTATGTAGAGAAAGCAAAGCAGAAAGGCGAGATAGACAGCCGAGAAAGTGCTTA 4406  
Db 144336 TCTACCTTAGATAGAGAGAGCAGACAGAGAAA-TACAGATAGAACAGTCAAAAGTCTTA 144394  
QY 4407 GTGTCCCTCCCAAGCGCTCTTCTAGATAGTGTGTGATCAGGTGTGCTGGTTGGTTGTTCA 4466  
Db 144395 GTGTCCCTCCCAAGCGCTCTTCTAGATAGTGTGTGATCAGGTGTGCTGGCG---TGCTTGTGTTCA 144451  
QY 4467 GTTAGCGTCACTCTGTGTCATGCTGAGGTAATGTGTGA---CCAGTCTTTTGTAGTGGTTA 4524  
Db 144452 GGTAGCTTCACTCTGTGTCATGTCAGCTCTTTAGTCCAGTCTTTACCTTGGGACTAAA 144511  
QY 4525 CATCCAAAAGAGAGACCTCTGAGTCGGGTGTG-GGATGAGCTTTTCCAGACCTG--- 4579  
Db 144512 CGTGCATTAAGAGAGCGCTCTGGATGTGTGTCGAGGATGAGCTTTTCTCCACAGAACT 144571  
QY 4580 GCAGGTAACACTCACTCAGTTTATAAATCTCCCTGGTTATTTCCGTTTGTATGATGATCTAAG 4639  
Db 144572 GGCTGGGTAACTCACTCAGTTTATAAATCTCC- -TGGTTATTCTGTCTGATGTGCTCTAAG 144629  
QY 4640 GTCTGCCCTCAGTGTG-ATGATGTTTATCCACACAGGTTAGTAAGAGTGCACAGCA 4698  
Db 144630 GTCTGCCCTCAGTGTGTCAGTGTATATCCACACAG---TAGAGAGTGCAGTACCA 144686  
QY 4699 GAAACGTTGGTCTTTATTTTGGAGAACCCCAATTTCTGTGTATTTTATGCACCTGCTTTTA 4758  
Db 144687 GAAACCTTGGGCTTTATTTTGGAGAACCTCATTTCTGTGAGCTTTTGTGCACCTGCTTTT 144746  
QY 4759 GTGAACCTCC-AGAGTGTGATTAAGAGTCTGTGTTAGTCCGTCGGATGGCTAGTTTA 4816  
Db 144747 GTGAACCTCCAGAGAGGGGCTTGCAAGCTCTGTGTTAGTGTATGAAATGGCCACAGAA 144806  
QY 4817 GAAGCTATGTTTGAAGAGCGCAAGTTGACTTTAGAGAAAGAGCTGCACAGTGTGTA 4876  
Db 144807 GCTGTGTTGTTGAAAGAGCGGCTGTGTGACTTTAGAGAAAGAACTGTGACATGGCTA 144866  
QY 4877 GACATTTCTTTT-AAACCGGACTGACGTTTAAACACACTTGTATTTTCAGATGATTTAGTTT 4935  
Db 144867 GACATTTCTTTTAAACCGAGTGCAGCTTAACACTGATTTTCAGAGGATTTAGTTGTTT 144926  
QY 4936 TTGTTTCTGAGACCCAGCACTGTATA----- 4962  
Db 144927 TGTTTCTGAGACCTAGCACCTGATTTTTCGAAGGACGGGATATCTATACATTGACGA 144986  
QY 4963 -----TTTAAATCTGTCAGATTAACCTTCACTATCAATAGCTAAATG 5009  
Db 144987 TGTGTTGAGATGTTTGAATAATCGTTCAGTTTACACTTCACTGTCAATAGAAATAATG 145046  
QY 5010 ACTCATGCTGCGACACATGTC-----CTGATGCTGCGCAAGAACAGAGGATCT----- 5057  
Db 145047 ACTGTTAAATAATACCATGCTGCGCAACAAATGTTGCGAAAGACAGAGGGTTTATATA 145106  
QY 5058 ---TTGACTGAAGGAGAAAA-----ACTGTATGTCATCCAGGCCCCAGAGAAAGAA 5107  
Db 145107 AATTGACTGAAGGAGAAAAAACAAGATTACTGTCTATTGTCTATCCCAATCCAG-GCAAGAA 145165  
QY 5108 CACTCCAGCGCAGCGCAGCGCAGCGATGTTGTTCTAGTTGAATACACATTTCA 5167  
Db 145166 CACTCAAGCGCAGCGCAGCGATGTTGTTCTAGCTG-----AATACTTTACACATGCA 145218  
QY 5168 AGTCTTGAGTGGTGTCTTT---AGATCTGTGTAGCATGTGAGGCTCTGTACAGGTGGGG- 5224  
Db 145219 GGTCTTACAGTGGTCTTTAGAGATCTGTGTAGCATGTGAGGCTCTGTACAGGTGGGGAC 145278  
QY 5225 CCACACTTCTGAGGGCTGAATATGGGCAACCCCTTTATCTAATCTTGAATCAAAACCGTCA 5284  
Db 145279 CCACACTTCTGAGTGTGAATATGGGCAAC-----CTAATCTTGAATCAAAACCGTCA 145332  
QY 5285 AATTATTTTATTTTAAATTAAGAAA-----GAGT 5314  
Db 145333 GATTATTTTATTTTAAATTTAAGAAAAGTAGAGATGGAATGAGTTGTTGTTTTTTTTTTT 145392

QY 5315 TGGGGAATGACATTTTGTGAGTTGGCTTTTCAGCTCAGTCATTTTACGTGTAAC---GTG 5372  
Db 145393 TTTGGAGTCAGATTTTGTGAGTTGGC---TTTCAGATTAGTCATTTTACATGTAACATAATG 145449  
QY 5373 GACATTTGTAGTAGCTCAGATTTATTTGTATATAATTTTAACTAACTGTGTAATTTGTAAT 5432  
Db 145450 TAGCTTTATAGCTCAGATTTACATTTGTATATAATTTTAACTAACTGTGTAATTTGTAAT 145509  
QY 5433 AAATATATTTTGCATTTTAAAA 5454  
Db 145510 AAATATGTTTGCATCATTAAAA 145531

AC094236 240836 bp DNA linear HTG 14-NOV-2002  
Rattus norvegicus clone CH230-3B24, \*\*\* SEQUENCING IN PROGRESS \*\*\*,  
4 unordered pieces.  
AC094236  
VERSION AC094236.5 GI:24958472  
HTG; HTGS PHASE1; HTGS DRAFT; HTGS\_ENRICHED.  
KEYWORDS Rattus norvegicus (Norway rat)  
SOURCE Rattus norvegicus  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
1 (bases 1 to 240836)  
REFERENCE  
AUTHORS Muzny, D., Marle, M., Metzker, M., Lee, S., Abramson, S., Adams, C., Alder, J.,  
Allen, C., Allen, H., Alebrooks, S., Amin, A., Anguiano, D.,  
Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,  
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,  
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,  
Bryant, N., Buhat, C., Burch, P., Burrell, K., Calderon, E.,  
Cardenas, V., Carter, K., Cavazos, I., Cesar, H., Center, A.,  
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,  
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,  
Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,  
Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,  
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,  
Egan, A., Escoto, M., Eugene, C., Evans, C.A., Falle, T., Fan, G.,  
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,  
Fraser, C.M., Gabidi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,  
Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,  
Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, C.,  
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,  
Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hoques, M.,  
Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,  
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,  
Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,  
Kowic, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,  
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,  
Lorensuhewa, L., Loulaeged, H., Lozano, R., Lu, X., Ma, J.,  
Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A.,  
Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,  
Mawhinney, S., McLeod, M.P., McNeill, T., Meenen, E.,  
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,  
Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,  
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,  
Nwaokemele, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K.,  
Pasternak, S., Paul, H., Perez, A., Perez, L., Primus, E., Pu, L.,  
Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.,  
Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R.,  
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,  
Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.,  
Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,  
Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajda, D.,  
Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sores, J.,  
Steinle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C.,  
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,  
Valas, R., Vera, V., Villaseana, D., Waldron, L., Walker, B., Wang, J.,  
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,  
Williams, G., Willson, R., Wlezyk, R., Wooden, H., Worley, K.,  
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,



Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstock, G. and Gibbs, R. A.

Direct Submission

Unpublished

2 (bases 1 to 240836)

Worley, K. C.

Direct Submission

Submitted (17-SEP-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 240836)

Rat Genome Sequencing Consortium.

Direct Submission

Submitted (14-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Nov 14, 2002 this sequence version replaced gi:22772949.

The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GAGP

Center clone name: CH230-3B24

----- Summary Statistics

Assembly program: Phrap; version 0.990329

Consensus quality: 194130 bases at least Q40

Consensus quality: 198891 bases at least Q30

Consensus quality: 202078 bases at least Q20

Estimated insert size: 194853; sum-of-contigs estimation

Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

-----

\* NOTE: Estimated insert size may differ from sequence length (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).

\* NOTE: This is a 'working draft' sequence. It currently consists of 4 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

\* 1 184581: contig of 184581 bp in length

\* 184582 184681: gap of unknown length

\* 184682 238285: contig of 53604 bp in length

\* 238286 238385: gap of unknown length

\* 238386 239477: contig of 1092 bp in length

\* 239478 239577: gap of unknown length

\* 239578 240836: contig of 1259 bp in length.

Location/Qualifiers

1. 240836

/organism="Rattus norvegicus"

/mol\_type="genomic DNA"

/db\_xref="taxon:10116"

/clone="CH230-3B24"

1984. . 3036

/note="wgs\_contig"

misc\_feature

FEATURES

source

ORIGIN

Query Match	Best Local Similarity	Score	2188.8;	DB 2;	Length	240836;
Matches	3655;	Conservative	0;	Mismatches	704;	Indels
55;						
Qy	791	CGCTCCACGGTGCAGAAAGTCTCTGAGTTACATCGCGCTGGACTCTTCGGAAGACTCG	850			
Db	174258	CGCTCCACGGTGCAGAAATCTTTGAGTT-CTCTGCGCTGGATTTTCTTCGGAAGACT	174200			
Qy	851	CGTCTGTCAGTACTCTTCGATTGGCTAGAGGAGGACTCCAGTGGCAGTGTCTGCTCCTCAG	910			
Db	174199	TGGTGTGTCAGTCTCTTGATCGGCTAGAGGAGGACTCCAGTGGCAGTGTCTGCTCCTCAG	174140			
Qy	911	ACTGGAAGTTAAACTCAAGGCCAGGAAAGAGCTTTAGACTCTGACGCGCCACTTTTCCT	970			
Db	174139	ACCTGGAATTTGAACTCAAGGCCAGGAAAGAGCTTTAGACTCTACAGCGCACACTTTCT	174080			
Qy	971	CCTGAGCAGCAGCTGTGGGAGTGGAGTTGCTGCCCGCTCTTCAAGCTGTCTAGT	1030			
Db	174079	CATGAGCAGCAGCTGTGGGAGTGGAGTTGCTGCCCGCTCTTCAAAACCGATCTAGT	174020			
Qy	1031	CTCCACCGGAGAACTTGACTCTTCTCTCTGCGCTCTCTGAGCGTTCTCAGAGCTTAGGTAA	1090			
Db	174019	CTCCACCAAGAACTTGGCTCTTCTCTCTGCGCTCTCTGAGCACTCAGAGCTTGGGCAA	173960			
Qy	1091	TTTCAAGGTAGTTTCTCTATCTCTGAAACCTTCTCTCTGAGCTACCTTCCCGAGTTAGG	1150			
Db	173959	TTTCAAGGTAGTTTCTCTATCTCTGAAACCTTCTCTCTGAGCTACCTTCCCGAGTTAGG	173900			
Qy	1151	GCTGCGCTCTCAGAGCAGCGCTGGAGTGGCGAGTTGTGGTTCGGAACACTTAACCCC	1210			
Db	173899	GCTGCGCTCTCAGAAACAGCATCTGAGTGGCCAACTTTGTGGATTTCCGAGCGCTATCCCC	173840			
Qy	1211	AGAGAGCTGCTATCTTCTGAAGATGTTGTCTCAACCTCTCAGCGTTCGGGGCAGAGATGTC	1270			
Db	173839	AGAGAGCTGTTTACTTCTTCTGAAGTGTGTGTCTCTCAGCGTTCGGGGCAGAGATGTC	173781			
Qy	1271	GGCAACCGCTCGAGAGAGTGTCCGCTCTCTCTACAGAGGCTTCGCGAAATCCACCA	1330			
Db	173780	GGCAACTGCTGCGGAGTGTCCAGCGCTCTCTACAGAGGCTTCGCGAAATCCACCA	173721			
Qy	1331	CCTTCGTATGAAACGGCTAGAAATCTCTCAGGCTTAACAAAGGCGAGAGTTACCCACC	1390			
Db	173720	CATTTCGATGAAACGGCTAGAAATCTCTCAGGCTTAACAAAGGCTTAACCCACC	173670			
Qy	1391	TGACCAAGATAATGGCTATCATAGCTTGAGAGGAGCAATAAATCTCTCCGGATGGACCC	1450			
Db	173669	TGACCAAGATAATGGCTATCATAGCTTGAGAGGAGCAATAAATCTCTCCGGATGGACCC	173610			
Qy	1451	ACAACATTGCACAGATAACCCAGCAGAGCGGTGTCCCTCTGTCGAGACAGCGCGAGCC	1510			
Db	173609	GCAACATTGCACAGATAAAGGTAGCAGAGCTATGTCCCTCTGTCGAGCCAGCGCAGCC	173550			
Qy	1511	CACTGAGAAAAACCAAGATTTGGTATTCAAGAGTTTCAAGAGCTTCACAGAGCCCGGAGCAG	1570			
Db	173549	CACTGAGAAAAACCAAGATTTGGTATTCAAGAGTTTCAAGAGCTTCACAGAGCCCGGAGCAG	173490			
Qy	1571	TCCTGTTTGTGAATTAACCCGTGAAATAAGATCTGAAGAGGACCACTAATCAACTGA	1630			
Db	173489	TCCTGTTTGTGAATTAACCCGTGAAATAAGATCTGAAGAGGACCACTAATCAACTGA	173430			
Qy	1631	CCTCTCAGATPAGAGGAGAGAGCTTCTCTGTTTCTTACAGACCAAGTTGTAGCAACAACT	1690			
Db	173429	CTACTCAGACATAGGGGAAACCTTCTCTGTTTCTTCCAGACCAAGTTGTAGCAACAACT	173370			
Qy	1691	GATAGATTATTTTGGAGGCGCCCGCAGTGTGATTTGGAAGCCAGCTCTGATTTCTGAAG	1750			
Db	173369	GATAGATTATTTTGGAGGCGCCCGCAGTGTGATTTGGAAGCCAGCTCTGATTTCTGAAG	173310			
Qy	1751	TGAGGATTTGGGCGCAGGAACCTCAGGACGATGCTTTGTATGAGGATGGCTCCCTGTCTGA	1810			
Db	173309	TGAGGATTTGGGCGCAGGAACCTCAGGATGATGATGTTTGTATGATGAGGCTCCCTGTCTGA	173250			
Qy	1811	ATCAGAGCTGGAAACAGGACTTCGGAAGGCGCTTCACTTTTGGAACTCTTTTCCACAGTGTAGA	1870			

Db	173249	ATCAGACGGGAAACAAGACTCGAAGGGCTTCACTTTGGAACTCTTTCTACAGTGTAGA	173190	Db	172177	TTTCTGTGTGCCACCATGTGCCAGCTGGAAGTGTGTTTTTAAACATCTTCTGTCTAACCTGGGT	172118
Qy	1871	TCCTTTACAAACCCCAAATTTTACAGCCACGATTCAGAGGCTGCCAGAAATGGCCCCAG	1930	Qy	2936	GA-----GAGTTTAAATTTTGCACCTTTCAAAAAGTCTTTCTTCTGGAGCAGGGC	2983
Db	173189	TCCTTTACAAACCCCAAATTTTACAGCCACGATTCAGAGTGTCTGCCAGAAATGGCCCCAG	173130	Db	172117	GAGAGGGGCCCTTTGAGTTTAAATTTCTGCACCTTTCAAAAATATTTCTTCTGGAGCAGAGC	172058
Qy	1931	AGACCCATCAGATTCAGGGACATCTGGTCTGGCAGCTGTGGTGTAGGGAGCTGTACGGA	1990	Qy	2984	TGCTGTGTTTTACATACAGGCTTCAGGCCAGCTAGGGCTATGTGGTGAGACCTAGTCTTA	3043
Db	173129	AGACCCATCAGATTCAGAGAGTCTGGTCTGGCAACCCGGATGTAGGGAGTCTCAGGC	173070	Db	172057	TGCTCAG-----CCAGCCGGGCTATGTGGTG-----AGTCCCTA	172024
Qy	1991	GGGACCCCTTCGGAGACCCCGACCATAGTTCGGGGAGGAAGATCACTGGGAACCGAG	2050	Qy	3044	GAGGACAAACAGAAACAAACAGTCAAGTCTGTGGAACCTGAGGAGGAGGATAGGAAG	3103
Db	173069	CACATCTCTTCCCGAGACCCCGACCATAGTTCGGGGAGGAAGATCACTGGGAACCGAG	173010	Db	172023	GAGGACAAACAGAAACAAACAGTCAAGTCTGTGGAACCTGAGGAGGAGGAGACAG	171965
Qy	2051	TGCAGATGAGACAGAGATCTTAAATTTGGAACCTTTTCTGTCAATCTGAGGACCCCTA	2110	Qy	3104	GTCCAGGTATGCTTGGGAACTCAGTGAACCTCAAGGCCAGCATGGGGAGTTTAGC	3163
Db	173009	TGCAGAGGAGCAGAAATCATAGATTTGTGGAATCTCTCTGTAAATTTCTGAGGACCCCTA	172950	Db	171965	TACCAGGTCTGCCTGGAATCTCAG-----GGAGCTCAAGGCCAGCGTGGGGCTTTAGT	171913
Qy	2111	CAACCTTTTAAATTTAAGGCTCTCTTTTCAACCGTCAAGGGAAGAAATTTGGAAGGCCCTCA	2170	Qy	3164	AGACCTTTTACACTCAGAAATGGA-----AAAGGAAGCTGGGGATATAGTTCGGTAGCAG	3219
Db	172949	CAACCTTTTAAATTTAAGGCTCTCTTTTCCAAACGTTAGGGAAGAGCTGGAAAGGCTGTGA	172890	Db	171912	GAGACCTTTTTCGTTTTGAGCTCAGAAATGGAACGGAGGCTGGGGATATAGCTCAGCAGCAG	171853
Qy	2171	GGACTCAAGGCCCTTCTGTAGGTACAGTGGCCCTTCTCTGGCCATCATACCTTTACTTTC	2230	Qy	3220	AGGCCCTGCTACATGTCAAGGTCTCTGGTTTCAGTCCCGAGTCTTGCAAAATAGAAAGA	3279
Db	172889	GGACTCGAAGGCATCTCTGGGGCCACAGTGGTCTCTGTCTGGGCATCAGACCTTTACTTTC	172830	Db	171852	AGGGCCCTGCTACATGTG--AAGGCCCTGGGGTCACTCCAGTCTTCCAGTCTTGCAGGAAGAGA	171794
Qy	2231	TTGTAAAGGCCCA-----GCTGTTTAGAGAGCAAGAGATAATTTGTCAGGCTGTGGCT	2284	Qy	3280	AAAAACAATTTGCTTGGGATAACTATAAGGTTTAAAGGCTTCATAGTCAGTCTTAACTCAAAAT	3339
Db	172829	TTGTAAAGGCACAGAACTTAATTTAGAGAGCAAGAAAGATTAATTTG--AGGCTGTGGCT	172771	Db	171793	AAA---ACAGTCTTGGATAACTATAAGCTTCAAGCTCATATTCAGTCTCTATCTCAGATT	171737
Qy	2285	GGGTGAGGCTCTTCTCGAGAAAGATACACCATATCAAGAGAAAGAAAGGTAACTTTCT	2344	Qy	3340	ATGCATGC-----ATTGAGTCTTGTGTGCTTTTTTCTGTTCTAAAAATTAATGGGC	3389
Db	172770	GGGTGAGGCTCTTTTGTGAAAGATACACCATATCAAGAGAAAGAAAGATAACTTTCT	172711	Db	171736	ATGCATACTCATATAAAACCTGAGTTCTTGTGTGCTTTTTTCTATATCAATAAATCAATGGT	171677
Qy	2345	GGAGAGGTTACTGAGTATTAATAAGTGTGTGATGAGATGCGAAAGCAATCGGGAAGA	2404	Qy	3390	TTTTATGGGTTT--TGTTTTGTATGTTTTATGCTGTATATTTTCAGGTAGGGGATTGA	3448
Db	172710	TGAAGAGATTAATGAGTATTAATAAGTGTGTGATGAGGCTCGCAAGGATTAATGGGAAGA	172651	Db	171676	TTTATGTGTTTTGTTTTCAATATTTGTGTGCTGTGTTTTATTTTGAAGCGGGGATTGA	171617
Qy	2405	ATTTGCAAGGAGTGAATGACGTTCCAGAAACGAATTAAGAAACAGAAAGTTCGATGG	2464	Qy	3449	GGCAGGCTCTCACTGTGGCTTCAGCTCATGGCAGTTTTCTGCTTCAACTGAGTGTGTTT	3508
Db	172650	ATTTGCAAGGAGTGAATGCAAGTTCCAGAAACGAATTAAGAAACAGAAAGTTCGATGG	172591	Db	171616	AGCAGGCTCTCACTGTGACCTT--GCTATGGCAGTCTTCTGCTTCAACCTCCTGAGTG	171559
Qy	2465	CTACTGTGGCTTTTGAGCAGAGAAATAATGTTTAAATAGACTGAGGATCGAGTCAAA	2524	Qy	3509	GGTTGTGTGTGTGTTTTTTTACTTTTCATAGATTTGACTTAATGAAGGCAAAACC	3568
Db	172590	TTACTGCTTGAGTTTGAGCACAAGAGTGTATAGTAGACTGCGAGACCGT-GTTTTCAA	172532	Db	171558	TGGCGCTATACTGTGTGTTTTTGTGTTTTGTTGTTTATGTGTTG-----	171513
Qy	2525	GGACTTACTGTGTACAGCAATGTTAA--GAAGTGAACAGCCTGCAACCCGTGCCACTC	2582	Qy	3569	TGTTATCAACTAAAGACATTCATGTGTCACCTTCAGTGGTGATTTTCTCCCTCTTTT	3628
Db	172531	GGACTTCTGTGTACAGCAGTGTCTAAACAGAGGGAGCAGCTGCAGCCCTCACCCACTC	172472	Db	171512	-----	171513
Qy	2583	TGCTCTTACTTGTAGAGT-TTCCCTTAAACAAACACTGGCAGCTGTCTTGGACATGT	2641	Qy	3629	TTTTTTCCCCCAGAGCTGAGGACCAAAACCAGGGCTTTTGCACTTGTAGGCAAGCGCTCT	3688
Db	172471	CATCTTACTTGTAGAGGGTTCTCTTATAGCAAAACACTGGCAGCTGTCTTGGACATGG	172412	Db	171512	-----	171513
Qy	2642	TTTTTAAAGAAACAATCTGTATCTAGAGATGAGTTTTGATATTTTGGGTAATGTGCTC	2701	Qy	3689	ACCAGTGAAGTAAATCCCCAAACCAACCCGCCCTTTCTTTTCTTTTAAAGACATGGTCT	3748
Db	172411	GTTTTAAGAAAGGACTTGGTTCTAGTGTAGTGTAGTTTATTTTGGATAATGTGCTC	172352	Db	171512	-----	171513
Qy	2702	ATTAGAAACAC---CAACTCCGATAATGAAGATCTCTTATCTGTATATCTCTCTTTTCC	2758	Qy	3749	TATATAGTCTAGGCTGGCTTTAACTCATTTACTTAGCCAAAGGATGGCTTTTGAAGATCTCT	3808
Db	172351	ATTAGAAACACACAACTCTAATAATGAGGAGTCTCTAATCTTTAATCTCTTTTCTTCC	172292	Db	171512	---GTAGTCTAGGCTGGCTTTGAACCTCATTTACTTAGCCAAAGGATGACTTGGAGAGATCTCT	171456
Qy	2759	TATTTAGTTGATGTTGGG---TTTGTCTTTTTCGAGGGTCTCACTGCATAATCTGTT	2815	Qy	3809	CCTGTCTCTGTGTTTTGGCTTGACATGGCATGTGCCATACATCCA-----GATTTTTCTC	3863
Db	172291	TATATAGTGTGATGTTGGGTTTTTTTTTTTTTGTAGAGGGTCTCACTGC-----TTGT	172238	Db	171455	CTTGTCTC-----TGGATTGCATGGCATGCACCATACATCCAGATTTGATTTTTTTC	171401
Qy	2816	TGGCTAGGCTGTATGTATGAGACAGGTTGGCCTTGAAGTTGTGTATTTCTTGCTCTG	2875	Qy	3864	TGTATCTAGTCTTTTTTAAGATATTTCCAAAGACTGCCCCCAAAATCCACAGTGCAGTACT	3923
Db	172237	TGGCTAGGCTGTATGTATGAGACAGGTTGGCCTTGAAGTTGTGTATTTGCGCTTTG	172178	Db	171400	TATATCTAGGCTATTTTACGATATTTCCAAAGGACTGCCCAAAACCCCAAAATACAGTACT	171341
Qy	2876	TTTCTGTGCCCAACATGCCAGCTGGAGTGTGTTTTTAAATATCTCTGCTTACTCTGGGT	2935	Qy	3924	TTCTTGACTGGGAAGGGGGTTGGTGTGCTTTTCAAGGCAACACCATTTATGCCAGC	3983
				Db	171340	TTCTAGACTGGGAAGCCGTTT-----TGCTTCTTAAAGGCAACGCGAGTGTGCCCCAGC	171285

```
Qy 3984 TGGTCTGCAAACT-----TAACATATTTGGTTTCTGTGCAAACTGTCCATTGGAA 4035
Db 171284 TAGGCTGCCAACTTTAAACATGTAGTATATTTGGTTCTGTGCAA---CGCAATTAGAA 171228
Qy 4036 GGTTCCTGTGTGTTGTAGTTTTCAGTTGAATGFGGCTCTCTGTGTAGCTTTCAAGAATGGTA 4095
Db 171227 GGCACCTGTGTGTGTAGCTCCAGTTGACTGGCTTTC---GTAGCTCTCAAGNAATGGCA 171171
Qy 4096 GAAATCATAAAGCACTCT--TAAGTAATCATTGCATTGTAGACATTTTTTTTTT----- 4146
Db 171170 AAGATTGTAAAGCACTCTGTGGGAAGTGATCATTTGCCATTGTAGCTTTTAAACTTT 171111
Qy 4147 -----TTTAACTAGGGGTAATTTTGGGAGACATGTGAGATATTCACATTTTGTGTTT 4194
Db 171110 AGAAGCCTAGAAATTTTCTAGGGGGTATCTAGGAGACATCTCAGAAA-----TCTGTTT 171057
Qy 4195 TATGTGCTTTAAACCAAGTTTACTTACACCTGTGCTCAGCACTGGCACTTTTCAAACCT 4254
Db 171056 TATGTATCTTAAACCAAGTTTACTTATACCTGTGCTCAGCAGGGGCACTTTTCAAACCT 170997
Qy 4255 GTCCTGGAGACACTGTAACTTGGATGGATGTCAGTTCTGGGTTTTCATGTGTAAATGTCA 4314
Db 170996 GTCCTGGGTGACCTGTAACTTGGATTTGTCAGTTTCTCAGTTTGCAGTATGNAATGTCA 170937
Qy 4315 ACTCAAAAGGTCAATATCCAGGTTTGTGTGTTCTTACCTTATGTAGAGAGCAAAAGCA 4374
Db 170936 AATACAAAGGTCA--ATCCAGGTCCATTGTGTTCTACCTTAGATAGAGAGCAAGCA 170878
Qy 4375 GAAAGGGCAGATPAGAGCAGCAAGAGTGTAGTGTCCCCCCCCAAAGCGTCTTCTAGAT 4434
Db 170877 GAAA--TACAGATAGAAACAGTCNCANAGTCTTAGTGTCTCCCGAAGCTTGTCTTGAAT 170819
Qy 4435 AGTGTGTGATCAGGTGTGCTTGTGTTTCTTGTAGTTAGGTCACCTGTGTGTCATGTTGA 4494
Db 170818 AGAGTGTGTGGCTTGGGCG---TGCTTGTTCAGGTAGTTCACCTCTGTGTCATGGGCAGC 170762
Qy 4495 GGTATATGTTTACCAGTTTCTTTAGTGGTTTACA--TGCAAAAAGAGAGGACCTCTGAGTC 4552
Db 170761 TCTTTAGTCCGACGCTTACCTGGGACTAAACGTGCACTAAGAGAGGCGCTCTGGATT 170702
Qy 4553 GGGTGTG--GGATGAGCTTTCCAGACCTG---GCAGGGTAAACTACCTCAGTTTATTAATC 4607
Db 170701 GGTCTGAGGATGAGCTTTTCTCAAGAACTGGCTGGGTAACCTACCTCAGTTTACAATC 170642
Qy 4608 TCCCTGTTATTTCCGTTTGTATGATCTAAGTCTGCTCAGTGTG--ATGATGTTTCAT 4666
Db 170641 TCC--TGTTTATCTGTCTGATGTCTTAAGGTCTGCCTCAGTGGTGTCAAGGTGTTAAT 170584
Qy 4667 CCACACAAAGTTAGTAAGAGTGCACGACAGAAACGTTGCTCTTATTTTGAAGAACCC 4726
Db 170583 CCACACACA--GTAGGAGGAGTGCGATACAGAAACCTTGGGCTTATTTTCGAGAACCC 170527
Qy 4727 CCATTTCTGTGTATTTATGACACCTGCTCTTTAGTGAATCC--AGAGTGCATTTAAAGAT 4784
Db 170526 TCAATTTCTGTGGACTTTGTGCACTGCTCTTTTGTGAATCCAGAGAGGCGCTTGCAAGT 170467
Qy 4785 CTGGTTTGTGCGCGGGAATGGGCTAGTTTAGAGCTATGTTTGGAAAGCAGGCGGT 4844
Db 170466 CTGGTTTGTGCTATGGAATATGGGCCACAGAAAGCTGTGTTTGGAAAGCAGGCGGT 170407
Qy 4845 GACTTTAGAGAAAGAGCTGTGACAGTGTGTAGACATTTCTTTT--AAACCGGACTCGAGC 4903
Db 170406 GACTTTAGAGAGAAAGCTGTGACATGCTAGACATTTCTTTTAAACACAGATGGCAGC 170347
Qy 4904 TTAACAAACATTGATTTTCAGATGATTAGTGTGTTTCTGAGACCCAGCACTGTATA- 4962
Db 170346 TTAACACTGATTTTCAGAGGATTAGTGTGTTTGTGTTTCTGAGACCTAGCACTGAAATAT 170287
Qy 4963 -----TTTAAAAATGTTCC 4977
Db 170286 TTGCAAGGAGCGGGATATCTATACATTCAGCATGTGTTTGATGATGTTTGAATATCGTTCC 170227
```

```
Qy 4978 AGATTACACCTTCACTATCAAAATGAGTAAATGACTCATGCTGCAGACATGTC----- 5030
Db 170226 AGTTTACACCTTCACTGTCAAAATGAAATGACTGGTGTAAATTAACCATGCTGCAGAA 170167
Qy 5031 CTGATGTGGCAAGAACAGAGGATCT-----TTGACTGAAGGAGAAAA--A 5075
Db 170166 CAAATGTGGCAAGACAGAGGGTTTTTATAAATTTGACTGAAGGAGAAAAACAAGTTA 170107
Qy 5076 CTGTCAATTTGTCATCCAGCCGCCAGGAAAGAACACCTCCAAAGGCGAGCGACGCGAGC 5135
Db 170106 CTGTCAATTTGTCAT--CCAATCCAGGAAAGAACACCTCAAGGCAAGGCGATGGCGGT 170048
Qy 5136 AGGCATGCTGTGTTCTAGTTTGAATACACATTCAGTCTTTGAGTGTGCTTT--AGATCTG 5193
Db 170047 TCTAGCTG-----AATACTTTACACATGACAGTCTTACAGTGTGCTTTAGAGATCTG 169995
Qy 5194 TGTAGCATGTGAGGCTCTGTACAGGTGGGG--GCCACACTCTTCTGAGGGCTGAATGTGGCA 5252
Db 169994 TGTAGCATGTGAGGCTCTGTACAGGTGGGGACCCACACTTCTGAGTCTGAAATGTGGCA 169935
Qy 5253 ACCCTTTATCTAACTTCGAATCAAAACCGTCAAAATTTTATTTTATAAATTTAAGAAA-- 5310
Db 169934 ACC-----CTAACTTGAATCAAAACCGTCAAAATTTTATTTTATAAATTTAAGAAA 169881
Qy 5311 -----GAGTTGGGGAATGACATTTTTTTGAGTTGGCCT 5342
Db 169880 TAGAGATGGAATGAGTTTGTTTTTTTTTTTTTTTTGGACTGAGATTTTTTGTGTTGCG-- 169823
Qy 5343 TTTGAGTCTGATCATTTTACGTGTAAAC--GTGAGATTTGATAGTCTCAGATATATTTGT 5400
Db 169822 -TTGAGATTAGTCATTTTACATGTAACATAATGAGGCTTTATAGCTCAGATTCATTTGT 169764
Qy 5401 ATATAATTTATTAATCTAATCTGTAATTTGTAATAATATATTTTGCATTTATTAA 5454
Db 169763 ATATAATTTATTAATCTAATCTGTAATTTGTAATAATATATTTTGCATTCATATA 169710
```

## RESULT 6

AC128396/c

LOCUS

DEFINITION

AC128396

AC128396.4

GI:23322373

HTG: HTGS PHASE1; HTGS DRAFT; HTGS\_ENRICHED.

Rattus norvegicus (Norway rat)

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

1 (bases 1 to 200490)

Muzny, D. Marie, Metzker, M. Lee, Abramson, S., Adams, C., Alder, J.,

Allen, C., Allen, H., Alsbrooks, S., Amin, A., Angiano, D.,

Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,

Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,

Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,

Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,

Cardenas, V., Carter, K., Cavazos, J., Ceasar, H., Center, A.,

Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,

Cleveland, S., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,

Devila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,

Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,

Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,

Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G.,

Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,

Fraser, C. M., Gabis, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,

Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,

Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K.,

Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,

Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M.,

Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,

Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,

Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,

AC128396 200490 bp DNA linear HTG 26-SEP-2002

Rattus norvegicus clone CH230-194B23, \*\*\* SEQUENCING IN PROGRESS

\*\*\*, 6 unordered pieces.

AC128396

AC128396.4 GI:23322373

HTG: HTGS PHASE1; HTGS DRAFT; HTGS\_ENRICHED.

Rattus norvegicus (Norway rat)

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

1 (bases 1 to 200490)

Muzny, D. Marie, Metzker, M. Lee, Abramson, S., Adams, C., Alder, J.,

Allen, C., Allen, H., Alsbrooks, S., Amin, A., Angiano, D.,

Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,

Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,

Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,

Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,

Cardenas, V., Carter, K., Cavazos, J., Ceasar, H., Center, A.,

Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,

Cleveland, S., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,

Devila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,

Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,

Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,

Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G.,

Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,

Fraser, C. M., Gabis, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,

Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,

Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K.,

Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,

Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M.,

Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,

Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,

Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,

Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorenshehwa, L., Loulseg, H., Lozano, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindartine, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhiney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwokeneme, O., Okwunonu, G., Olarunpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, P., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sleson, I., Sitter, C.D., Smajs, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villagana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausen, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

Direct Submission  
Unpublished  
2 (bases 1 to 200490)  
Worley, K.C.

Direct Submission  
Submitted (19-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 200490)  
Rat Genome Sequencing Consortium.

Direct Submission  
Submitted (26-SEP-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Sep 26, 2002 this sequence version replaced gi:22771040.  
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). As a result, the sequence may extend beyond the ends of the clone and there may be contigs that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: GYZL  
Center clone name: CH230-194B23  
----- Summary Statistics  
Assembly program: Phrap; version 0.990329  
Consensus quality: 131649 bases at least Q40  
Consensus quality: 136220 bases at least Q30  
Consensus quality: 138894 bases at least Q20  
Estimated insert size: 161200; sum-of-contigs estimation  
Quality coverage: 4x in Q20 bases; sum-of-contigs estimation

-----  
\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
\* NOTE: This is a "working draft" sequence. It currently  
\* consists of 6 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence

\* as soon as it is available and the accession number will  
\* be preserved.

1 21048: contig of 21048 bp in length  
\* 21049 21148: gap of unknown length  
\* 21149 108746: contig of 87598 bp in length  
\* 108747 108846: gap of unknown length  
\* 108847 161819: contig of 52973 bp in length  
\* 161820 161919: gap of unknown length  
\* 161920 176640: contig of 14721 bp in length  
\* 176641 176740: gap of unknown length  
\* 176741 187347: contig of 10607 bp in length  
\* 187348 187447: gap of unknown length  
\* 187448 200490: contig of 13043 bp in length.

FEATURES  
source  
1..200490  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clone="CH230-194B23"  
1..1226  
/notes="wgs contig"  
misc\_feature  
13967..15951  
/notes="wgs contig"  
misc\_feature  
19893..21048  
/notes="wgs contig"  
misc\_feature  
21149..23404  
/notes="wgs contig"  
misc\_feature  
29744..33119  
/notes="wgs contig"  
misc\_feature  
48098..50488  
/notes="wgs contig"  
misc\_feature  
71079..72931  
/notes="wgs contig"  
misc\_feature  
72982..74735  
/notes="wgs contig"  
misc\_feature  
89602..91141  
/notes="wgs contig"  
misc\_feature  
176741..177999  
/notes="wgs contig"  
misc\_feature  
185991..187347  
/notes="wgs contig"  
misc\_feature  
187448..188936  
/notes="wgs contig"

ORIGIN  
Query Match 30.9%; Score 1689.2; DB 2; Length 200490;  
Best Local Similarity 84.5%; Pred. No. 0;  
Matches 2022; Conservative 0; Mismatches 323; Indels 49; Gaps 9;  
Qy 1 CGGTCTCTCGCTCTCGCCCTTCGACGTTCCGGGTGTGGGCTGCGGCCATTTTGAGCTTCG 60  
Db 197561 CGGTCTCTCGCTCTCGCTTCGTTGCGAGCTTCGGGTTCGGCTGCGGCCATTTTGAGCTTCG 197502  
Qy 61 TTCTTTGGCCCTCGCTGCGCCACCCAGCCACCCCTTTCCGCTTGGCGTTTCGGCCCTCCG 120  
Db 197501 TTCTTTGGCTCTCGCCCGCCACCCAG----CCCTAAAGCTGTGTATTTTCGCGCTCCG 197446  
Qy 121 TGGGGCCACCGGAAACGCGCGCTGTCTCCGTCGCGCGCGCGGAGGAGGTCTTCT 180  
Db 197445 TGTGGGCGCG-----GGTCTCCGAAAGCTTCCGCGCGCGCGGTGTGAGGGGTCTTCT 197392  
Qy 181 CTATGTTGGAGCGATCTCACACGGCTTAGAGCGTCTCTTCTCCCTAGCCGGATGGACCTA 240  
Db 197391 CTATGTTGGAGCGGTCTCACACGACTTAGGGGCGCTCCAT-TCTCCCGAGATGGAGCTA 197333  
Qy 241 ACCGCGTTCGCCACCGCTTGGCGGGCGCTTGGGCGCTCGGTGCGAGCACTCGTTGGGA 300  
Db 197332 GTCCGCTCGCCACCGCTTGGCGGAGCGCGCGCGGTACCGTACATCGCTCGCTACGGA 197273  
Qy 301 AGCGCGGCTCTCTGGGCGCTCTCTCGCGCGGGAAATCGGACTGCGACTGCGACTCCG 360  
Db 197272 AGCGCGGCTCTCTGGG-----CCTCTGTCGCGGGAAATCGGATGCGACTCCG 197216

QY 361 TGGCTGGCAAGCGGAGACTGTGTAGACCTCGGATCCAGCTCGGCTGACGCGCGTGGAG 420  
 Db 197215 TGTCTGGGCAAGGCTAAGACTGTGTAGGCTCGGATCCTGGCGCTCTGAACGCGCTGAG 197156  
 QY 421 CTCTGTCTCTCTCTGTCTGAGAACGCGCAAGGAAGAGATGAGAGACAGGAACGACCA 480  
 Db 197155 CTC---CGTACTCGTGGCCGGAAGTTCCCAAGGAAGAGATGAGAGACGGGAACACACA 197099  
 QY 481 GGGCCGGAAGCGGCTCGGCCCTCGGCTCGGCTCGGCTCGGCTCGGCTCGGCTCGGCTCGG 540  
 Db 197098 GGGCTCGGAAGCGGCTCGGCTCGGCTCGGCTCGGCTCGGCTCGGCTCGGCTCGGCTCGG 197039  
 QY 541 GATCGCACGCTGTCTTTCGAGTTCCTCGGCTCGGCTCGGCTCGGCTCGGCTCGGCTCGG 600  
 Db 197038 GTTTCGACGCTGTCTTTCGAGTTCCTCGGCTCGGCTCGGCTCGGCTCGGCTCGGCTCGG 196979  
 QY 601 CCGCTCTCGGCGGCTCGGACGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAG 660  
 Db 196978 GTTTCGACGCTGTCTTTCGAGTTCCTCGGCTCGGCTCGGCTCGGCTCGGCTCGGCTCGG 196919  
 QY 661 TGCTCCCTAGCTATTTCCAGAAAGCTGTCTTTCGAGTCTGAGTCTGAGTCTGAGTCTG 720  
 Db 196918 TGCTCCCTAGCTATTTCCAGAAAGCTGTCTTTCGAGTCTGAGTCTGAGTCTGAGTCTG 196859  
 QY 721 CTAACAGATGGCTAGATTTTCGCAAGTTCAGCGCTCTGAGTCTGAGTCTGAGTCTGAG 780  
 Db 196858 CGACAGATGGTGTAGATTTTCGCAAGTTCAGCGCTCTGAGTCTGAGTCTGAGTCTGAGTCTG 196799  
 QY 781 AGGAATCTGACGCTCCGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTG 840  
 Db 196798 AGGAATCTGACGCTCCGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAG 196740  
 QY 841 CGAAGACTCGGCTCGTCAATTTTCGCAAGTTCAGCGCTCTGAGTCTGAGTCTGAGTCTG 900  
 Db 196739 TCGGAAGACTTGGTGTGAGTCTCTTTCGCAAGTTCAGCGCTCTGAGTCTGAGTCTGAG 196680  
 QY 901 TCGTCTCTCAGA-CTGGAAGTTAAATCTCAAGGCCAGGAAGAGCTTTTGAATCTCTGCA 959  
 Db 196679 TCGTCTCTCAGACCTGGAATTTGAATTTCAAGGCCAGGAAGAGCTTTTGAATCTCTG 196620  
 QY 960 CCACTTTCTCTGAGCAGGCTGTGGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGG 1019  
 Db 196619 CACACTTTCTCTGAGCAGGCTGTGGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGG 196560  
 QY 1020 GCTGGTCTAGTCTCCCAAGGAACTTGACTCTTCACTCTCTGAGGCTCTGAGGCTCTGAG 1079  
 Db 196559 ACCGGTCTAGTCTCCCAAGGAACTTGACTCTTCACTCTCTGAGGCTCTGAGGCTCTG 196500  
 QY 1080 AGCTTAGGTAATTTCAAGTATGTTCTCTATCTCTGAAACCTTCTCTACCTGGAATCACTT 1139  
 Db 196499 AGCTTGGGCAATTTCAAGTATGTTCTCTATCTCTGAAACCTTCTCTACCTGGAATCA 196440  
 QY 1140 CCCCAGTTAGGGCTGCGCTCTGAGCAGCAGCTCTGAGGCTGGAGTGGAGTGGAGTGGAGTGG 1199  
 Db 196439 CCCCAGTTAGAGCTGCGCTCTGAGCAGCAGCTCTGAGGCTGGAGTGGAGTGGAGTGGAG 196380  
 QY 1200 ACCTTAACCCAGAGAGCTGTCTCTTCTGAAGATGTTCTGACCTCTGAGGCTGGAGTGG 1259  
 Db 196379 ACATATCTCCAGAGAGCTGTCTCTTCTGAAGATGTTCTGACCTCTGAGGCTGGAGTGG 196320  
 QY 1260 GCAGAGATGTCCGGCAACCGCTCGGAGAGAGGTGTCCGCTCTCTCTACAGAGAGGCTGCG 1319  
 Db 196319 GCTGAGATGTCCGGCAACCGCTCGGAGAGGTGTCCGCTCTCTCTACAGAGAGGCTGCGCA 196260  
 QY 1320 GAATTCACACCTTCGTATGAACGGCTAGAAATTCCTCGAGGCTTAACAAAGGCGCAGAG 1379  
 Db 196259 GAATTCACACCTTCGTATGAACGGCTAGAAATTCCTCGAGGCTTAACAAAGGCGCAGAG 196200  
 QY 1380 TTACCCACCTCTGACCAAGATAATGGCTATCATAGCTTGGAGGAGGAACATACTCTCTC 1439  
 Db 196199 TTACCCACCTCTGACCAAGATAATGGCTATCATAGCTTGGAGGAGGAGCACTGCTCTC 196140  
 QY 1440 CGBATGGACCCCAACAACATTTGCACAGATAACCCAGCAGAGCGGTGTGCCCTGTCTGCAGAC 1499

Db 196139 CGGATGATCCGCAACATTTGCACAGATAAGGCGACAGGCAAGTCTCCCTCTCTGAGGCC 196080  
 QY 1500 AGGCGGAGGCCACTGAGGAAGAACCAAGATTTGGTGAATTCAGAAAGTTTTCACAGAGCCCC 1559  
 Db 196079 AGCCCGAGGCCACTGAGGAAGAAAGTGAATTTGGTGTGTAGGAAGTTTTCACAGAGCCCC 196020  
 QY 1560 CAGGGAAGCAGTCTGTCTTGTGAATTTACCCGTGGAAAAAGAAATGTGAAGAGGACCAACT 1619  
 Db 196019 CAGGGAAGCAGTCTGTCTTGTGAATTTAGTGTGCAAAAGAAATGTGAA----- 195972  
 QY 1620 AATGCAACTGACCTCTCAGATAGAGAGAGAGCCTTCTGTCTTCTACAGACCAAGTTTGT 1679  
 Db 195971 -----GACTACTCAGACATAGGGGTAACCTTCCAGTCTCTACAGACCAAGTCTGT 195921  
 QY 1680 AGCAACAACTGATAGATTAATTTTGGGAGGCGCCCCCAGTGAATTTGGAAGCCAGCTCT 1739  
 Db 195920 ACCAACAACTGATAGATTAATTTTGGGAGGAGCTTCCAGTGAATTTGGACACCAAGCTCC 195861  
 QY 1740 GATTCTGAAAGTGAAGTTCGGGCGAGGAACCTTGAGGACGATGGCTTTGATAGCGATGGC 1799  
 Db 195860 GATTCTGAAAGTGAAGTTCGGGATGAGGAACCTTGAGGATGATGGTTTTGATGATGATGGC 195801  
 QY 1800 TCCCTGTCTGAATCAGACGCTGGAACAGGAATCGGAAGGCCCTTCACTTTGGAACCTTTTC 1859  
 Db 195800 TCCCTATCTGAATCAGACGAGGAACAGGAATCCGGAAGGCTTCACTTTTGGAACTTTTC 195741  
 QY 1860 CACAGTCTGATCTCTTCAAAACCCCAAACTTTTACAGCCACGATTCAGAGCGCTGCCAGA 1919  
 Db 195740 TACAGTGTAGATCTCTTCAAAACCCCAAGACTTTTACAGCCAAATTCAGATCTGTGCCAGA 195681  
 QY 1920 ATTGCCCCCAGAGACCCATCAGATTCAGGGGACATCTCTGGTCTTGGCAGCTGTGGTGTAGGG 1979  
 Db 195680 ATTGCCCCCAGAGACCCATCTGATTCAGAGAGTCTCTGGTCTGGCACTCTGATGTAGGG 195621  
 QY 1980 AGCTGTCAAGAGGGAACCCCTTCCGAGAGACCCCGACCAATAGTTCGGGGAGGAAGATGAC 2039  
 Db 195620 AGTTCTCAGGCCACACTCTCTTCCCGACCCCTGACCAAGTTCCTGGGGAGGAGGATGAC 195561  
 QY 2040 TGGGACCGAGTGCAGATGAAGCAGAGAACTTTAAATTTGGAACCTCTTTCTGTCTATTCT 2099  
 Db 195560 TGGGAATCCAGTGCAGATGAAGCAGAAATCTTAGATTTGGAACCTCTTTCTGTCTATTCT 195501  
 QY 2100 GAGGACCCCTACAACTTTTAAATTTTAAAGGCTCTTTTCAACCGTCAGGGAAGAAATTGG 2159  
 Db 195500 GAGGACCCCTACAACTTTTAAATTTTAAAGGCTCTTTTCAAAAGCTCAGGGAAGAGCTGG 195441  
 QY 2160 AAAGGCGCTCAGGACTCAAGGCTCTTCTGAGGTCAAGTGGCTTCTCTGGCCATCAT 2219  
 Db 195440 AAAGGCGATC-----AGGCATCTCTGAGGCCACAGTGGTCTTTTCTGGGCAATCAG 195390  
 QY 2220 ACCTTACTTTTGTAAAGGCCAGCTGTTAGAGAGCCCAAGAGATTAATTTGTCAGAGGCTGT 2279  
 Db 195389 ACCTTACTTTTGTAGGCACTGTTAGAGAGCCCAAGAGATTAATTTGTCAGGCTGT 195330  
 QY 2280 GGGCTGGGTGAGGCTCTTGTGGAAGAAATACACCCATATCAAGAGAAAAAGGTAAAC 2339  
 Db 195329 GGGCTGGGTGAGGCTCTTGTGGAAGAAATACACCCATATCAAGAGAAAAAGGTAAAC 195270  
 QY 2340 TTCTTGAAGAAGTTACTGATTAATTAAGTGGTGTGATGAGGATCCCAAGGA 2393  
 Db 195269 TTCCNTTACATGACTTAAATTTGTTGTTATGTTTACAGTAGAAGA 195216

RESULT 7  
 AC098956/c  
 LOCUS Rattus norvegicus clone CH230-191M14, WORKING DRAFT SEQUENCE, 3  
 DEFINITION unordered pieces.  
 ACCESSION AC098956  
 VERSION AC098956.6 GI:30581544  
 KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.  
 SOURCE Rattus norvegicus (Norway rat)





QY 301 AGCGCGCTCTCTGGGCTCTCTGCGCGCGCGGAAATCGAGTACCGACTCCG 360  
 Db 152764 AGCGCGCTCTCTGGG---CCTCTGCTGGCGCGGAAATCGAATGCAATACCGACTCCG 152708  
 QY 361 TGGCTGGGCAAGCGGAGACTGTGTAGACCTCGGATCCAGCTCGGCTGGAGCGCGCTGAG 420  
 Db 152707 TGTCTGGGCAAGGCTTAAGACTGTGTAGGCTCGGATCCTGCGCGCTCTGAGCGCGCTGAG 152648  
 QY 421 CTCTGTCTCTCTCTCTGCTGAGAGCGCGCAAGGAAAGGAGATGAGACAGGAAGCACA 480  
 Db 152647 CTC---CCTACTCTGCTGCCGAGAGAGTTCGCAAGGAAAGGAGATGAGACGCGGAAACACA 152591  
 QY 481 GGGCCCGGAAGCGGCTCGGCCCTCGGCTGGGCTCTCTGCTGCGGCTGCGCTCTCTCTCGGC 540  
 Db 152590 GGGCTCGGAAGCGGCTACCTCTCGGATGGGCTCTCTGCTGCGGCTGCGCTCTCTCTCGGC 152531  
 QY 541 GATCGCAGCGCTGCTCTCTCGAGTTCGCGGCTTCCGCGCGCTTCTCTCTCGACAAATCCCGGAACT 600  
 Db 152530 GTTCGCACGCGCTGCACTTCGAGTTCGCGAAGCGCTTCTTCTCGACCAATCTCGAGGTCT 152471  
 QY 601 CGGCTCTGCCGAGCGTCCGACACAGGTACTGACCAAAATGCTTCTCTCAGCTCTCTTGCGC 660  
 Db 152470 CCGCGCGCGCGGCTCCGACAGGTACTGACCAAAATGCTTCTCTCAGCTCTCTTGCGC 152411  
 QY 661 TGCTCCCTAGCTATTCAGAAAGCTGCTGCTTTGGAGCCAGCTTTTCCGGGGGCTGATTC 720  
 Db 152410 TGCTCCCTAGCTTGCTCCAGAAAGCTGCTTAATTTGGAGTCACTTTTGGTGGCATGATTC 152351  
 QY 721 CTACAGATGGCTAGATTTTCCGCGAAGTTACAGGCGCTGAGAGCTTCGAGAGGACGGG 780  
 Db 152350 CGACAGATGGTATGATTTTTCAGCAAGTTACAGCGCTTAAGAGCTTTGAGAGGACGGG 152291  
 QY 781 AGGAATCTGACGCTCCACGCTGCAGAGTCTCTGAGTTACACTCGGCTCGACTCTTCG 840  
 Db 152290 AGGAATCAGCGCTCCACAGTGCAGAAATCTTTGAATTCCTGCGCTGACTCTTCT 152232  
 QY 841 CGAAGACTCGCGTCTGACTACTCTTGCAATGGCTAGAGGAGGACTCCAGTGGCAGTGC 900  
 Db 152231 TCGGAAGACTTGGTTGTGCTGCTCTTGTATGGCTAGAGGAGGACTCCAGGGGCACTGC 152172  
 QY 901 TCGTCTCTCAGA-CTGGAAGTTAAACTCAAGGCCCAAGAAAGCTTTAGACTCTGACGCG 959  
 Db 152171 TCGTCTCTCAGACTGGAATGGAATTCAGGCCCAAGAAAGCTTTAGACTCTACAGCG 152112  
 QY 960 CCCACTTCTCTCTGAGACAGCTGTGGGAGTGGAGTTGCTGCCAGTAGCCTTCAA 1019  
 Db 152111 CACACTTCTCTCTGAGACAGCTGTGGGAGTGGAGTTGCTGCCCGTAGTCTTCAA 152052  
 QY 1020 GCTGGTCTAGTCTCCCAACCGAGAACTTGACTCTTCACTCTCTGCGGCTCTGAGCGTTGAG 1079  
 Db 152051 ACCGGTCTAGTCTCCCAACCGAGAACTTGGCTCTTCACTCTCTGCGGCTCTGAGTACTCAG 151992  
 QY 1080 AGCTTAGGTAAATTCAGGTAGTTCTCTATCTCTGAAACCTTCTCTACCTGGACTACCTT 1139  
 Db 151991 AGCTTGGGCAATTTCAAGGTAGTTCTCTATCTCTGAAACCTTCTCTACCTGGACTGCTT 151932  
 QY 1140 CCCAGTTAGGGCTGCGCTGTGAGACAGCTGAGGTGGCCAGTTTGTGGTTTCCGA 1199  
 Db 151931 CCCAGTTAGAGCTGCGCTGTGAGAACAGCTGAGGTGGCCAACTTGTGGATTTCCGA 151872  
 QY 1200 ACATAACCCAGAGAGCTGTATCTTTCTGAAGATGGTTGTCAACCTCAGCGGTGCGG 1259  
 Db 151871 ACATAATCCAGAGAGCTGTACTCTTCTGAAGATGCTTGTCAACCTCAGCGCTGTGT 151812  
 QY 1260 GCAGAGATGTCGCAACCGCTGAGAGGTGTCGCTCTCTCTACAGAGGCTGCGG 1319  
 Db 151811 GCTGAGATGTCGCAACCTGCTGCGAGGTGTCGCGCTCTCTCTACAGAGGCTGCGCA 151752  
 QY 1320 GAAATCCACCACTTCTGATGAAACGGCTAGAAATTCCTCCAGGCTTAACAAAGGCGCAAG 1379  
 Db 151751 GAAATCCACCACTTCTGATGAAACGACTAGAAATTCCTTCCAGGCAACAAAGGCGCAAG 151692  
 QY 1380 TTAACCCACCTGACCAAGATAATGGCTATCATAGCTTGGAGGAGGAACATAACCTTCTC 1439

Db 151691 TTACCCACCCCTGACCAAGACAATGGCTATCATAGCTTGGAGGAGGACATGCTCTTCTC 151632  
 QY 1440 CGGATGAGACCCACAACATTGCAGATAACCCAGCACAGCGGTGTCCTCTGCTGCAGAC 1499  
 Db 151631 CGGATGATTCGCAACAAATTCACAGATAAGGACAGCAGCTCTCCCTCTCTGAGACC 151572  
 QY 1500 AGGCGGAGCCCACTGAGAAAAAACCAAGAAATGGTGATTCAGAAAGTTTTCACAGAGCCCC 1559  
 Db 151571 AGCCGAGCCCACTGAGAAAAAAGTAGAATGGTGGTTGAGGAAGTTTTCACAGAGCCCC 151512  
 QY 1560 CAGGGAAGCAGTCTGTTTTGTGAATTAACCGGTGGAAGAAAGATGTGAAGAGGACCACT 1619  
 Db 151511 CAGGGAAGCAGTCTGTTCTGTGATTAAGTTGCAAAAAGAAATGTGAA----- 151464  
 QY 1620 AATGCAACTGACCTCTCAGATAGAGGAGAGAGCTTCTCTTCTTCTACAGACCAAGTTTGT 1679  
 Db 151463 -----GACTACTCAGACATAGGGGGTAACTTCCAGTCTCTACAGACCAAGCTTGT 151413  
 QY 1680 AGCAACAAAATGATAGATTAATTTTGGGAGGGGCCCCCAGTGACTTTGGAAGCCAGCTCT 1739  
 Db 151412 ACCAACAAAATGATAGATTAATTTTGGGAGGAGCTTCCAGTGACTTTGGACACCACTCC 151353  
 QY 1740 GATTCGAAAGTGAAGATTGGGGGAGGAACTCTGAGACGATGGCTTTGATAGGATGGC 1799  
 Db 151352 GATTCGAAAGTGAAGATTGGGATGAGGAACTCTGAGGATGATGGTTTGTAGTGTAGTGGC 151293  
 QY 1800 TCCTGTCTCAATCAGACGTTGGAAACAGGACTCCGGAAGGCTTCCACCTTTGGAACCTTTTC 1859  
 Db 151292 TCCCTATCTCAATCAGACGTTGGAAACAGGAACTCCGAGGGCTTCCACCTTTGGAACCTTTTC 151233  
 QY 1860 CACAGTGTAGATCTTCAAAACCCCAAAATTTTACAGCCACGATTCAGACGGCTGCCAGA 1919  
 Db 151232 TACAGTGTAGATCTTCAACACCCCAAGAACTTTTACAGCCACAAATTCAGACTGCTGCCAGA 151173  
 QY 1920 ATTGCCCCCAGAGACCCATCAGATTTCAGGACATCTCTGCTGTGCGAGCTGTGTGTAGGG 1979  
 Db 151172 ATTGCCCCCAGAGACCCATCTGATTCAGAGAACTCTCTGCTGTGCGAACTCTGATGTAGGG 151113  
 QY 1980 AGCTGTCCAGAGGAGCCCTTCCGAGACCCCGACCATAGTTCCGGGGAGGAAGATGAC 2039  
 Db 151112 AGTTCTCAGGCCACATCTCTTCCAGACCCCTGACACAGTTTCCGGGGAGGAGGATGAC 151053  
 QY 2040 TGGGAACCGAGTGCAGATGAAGCAGAGAAATCTTAAATTTGTGGAATCTTTCTGTCAATCT 2099  
 Db 151052 TGGGAATCCAGTGCAGATGAAGCAGAAAACTTTAGATTGTGGAATCTTTCTGTAAATCT 150993  
 QY 2100 GAGACCCCTACAACTTTTAAATTTTAAAGCTCTCTTTTCAACCGTCAAGGGAAGAAATGG 2159  
 Db 150992 GAGACCCCTACAACTTTTAAATTTTAAAGCTCTCTTTTCAAAACGTCAGGGGAAGCTGG 150933  
 QY 2160 AAAGGCGTCCAGGACTCAAAAGGCTCTTCTGAGGTCAAGTGGCTTCTCTGGGCATCAT 2219  
 Db 150932 AAAGGCGATC-----AGGCATCTCTGAGGCCACAGTGTCTTTCTGGGCATCAG 150882  
 QY 2220 ACCTTACTTTCTTGTAGGCCACAGCTGTTAGAGAGCCAAAGATATAATGTCTCAGGCTGT 2279  
 Db 150881 ACCTTACTTTCTTGTAGGCCACAACTGTTAGAGAGCCAAAGACAAATTTGTCCAGGCTGT 150822  
 QY 2280 GGGCTGGGTAGGCTCTTGTCTGAGAAAGATACACCATATCAAGAGAAAAAGGTAAACC 2339  
 Db 150821 GGGCTGGGTAGGCTCTTGTCTGAGAAAGATACACCATATCAAGAGAAAAAGGTGTTGT 150762  
 QY 2340 TTCTCTGGAAGAAGTTACTGAGTATTAATAAGTGGTGTGATGAGGATCCCAAGGA 2393  
 Db 150761 TTCCATTTACTGACTATTAATTTGTGATTTTGTATGTTGATGTTACAGTAGAGA 150708

RESULT 8  
 AC105827  
 LOCUS  
 DEFINITION Rattus norvegicus clone CH230-101J7, \*\*\* SEQUENCING IN PROGRESS \*\*\*.

AC105827 248458 bp DNA linear HTG 21-SEP-2002  
 Rattus norvegicus clone CH230-101J7, \*\*\* SEQUENCING IN PROGRESS \*\*\*.



AC105927	AC105927.3	GI:23265261	HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ENRICHED.
VERSION	HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ENRICHED.		
KEYWORDS	Rattus norvegicus (Norway rat)		
SOURCE	Rattus norvegicus		
ORGANISM	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.		
REFERENCE	1 (bases 1 to 248458)		
AUTHORS	Mazny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J., Allen,C., Allen,H., Albrooks,S., Amin,A., Anguiano,D., Anyelebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,P., Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,K., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrill,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,M., Gunaratne,P., Haaland,W., Hamill,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Herrandes,B., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,L., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhea,L., Loulsegged,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhney,S., McGeed,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaokemele,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Polindexter,A., Popovic,D., Primus,E., Pu,L.-L., Puzo,M., Quiroz,J., Rachlin,B., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shivartsbeyn,A., Sibson,I., Sitter,C.D., Smajda,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Taber,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villaseana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,P., Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Zhao,S., Dunn,D., von Niederhausen,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstein,G. and Gibbs,R.A.		
REFERENCE	2 (bases 1 to 248458)		
AUTHORS	Worley,K.C.		
TITLE	Direct Submission		
JOURNAL	Submitted (10-JAN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA		
REFERENCE	3 (bases 1 to 248458)		
AUTHORS	Rat Genome Sequencing Consortium.		
TITLE	Direct Submission		
JOURNAL	Submitted (21-SEP-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA		
COMMENT	On Sep 21, 2002 this sequence version replaced gi:21736987.		

Qy	241	ACCGGGTCGCCACCGCTTGCGCGGGCCCTCTGGGCGCTCGGTGACGACTCTGTTGGCGA	300
Db	4175	GTCCGGGTGCGCACCGCTTGCGCGAGCGCCGCGCGGTACGGTACATCGCTCGCTACGCGA	4234
Qy	301	AGCGCGCGCTCTCTGGGCGCTCCTCTGCGCGCGCGGAATCGCATGTCAGTACCCACTCCG	360
Db	4235	AGCGCGCGCTCTCTGGG---CCTCTGCTGGCGCGGGAAATCGGAATGCAGTACCCACTCCG	4291
Qy	361	TGGCTGGGCAAGGCGGAGACTGTGTAGACCTTCGGATTCAGCCTCGGTGTAGCGCGCTGAG	420
Db	4292	TGTCTGGGCAAGGCTTAAGACTGTGTAGGCGCTTCGGATCTCTGCCCGCTCTGACCGCGCTGAG	4351
Qy	421	CTCTGTCTCTCTCTCTGCTGTAGAAAGCGCCCAAGGAAGGAGATCGAGACGGAACGCACA	480
Db	4352	CTC---CGTACTCTGTGCCGAGAAAGTTGCCAAGGAAGGAGATGGAGACGGGAAACACACA	4408
Qy	481	GGGCGCGGAAGCGGCTTGCCCTCGGCTGGGCTCTCTGTTCCGGCTGCCCCCTCTCTCGGC	540
Db	4409	GGGCTCGGAAGCGGCTTACCCTCTCGGATGGGCTCTCTGTTCCGGCTGCCCCCTCTCTCGGC	4468
Qy	541	GATCGCAGCGCTCTCTTCGAGATTCCGCGCGCCTTCTCTCTCGACAAATCCCGGGAACT	600
Db	4469	GTTTCGCACGCTGCACCTTTCGAGTTCGGAAGCCCTTCTTCGACCAAAATCTCGAGTCT	4528
Qy	601	CCGCTCTGCCCGAGCGTTCGGAACCAAGTACTGGAACAAATTGCTTTCTCAGCTCTCTGCC	660
Db	4529	CCGCGCGCGCGAGCCTCGGACCAAGTACTGGAACAAATTGCTTTCCGAGCTCTCTGCC	4588
Qy	661	TGTCCTCTPAGCCTATTCACGAAGCTGTCTTTGGAGCAGCTTTTCGGGGGCTCTGATTC	720
Db	4589	TGCTCCTCTAGCTTGCTCCTCAGAAGCTGTCTTATTTTGGAGTCAGCTTTTCGGTGGCATGATTC	4648
Qy	721	CTACCAAGATGCTAGATTTTTCGCCCAAGTTACAGCGCCTCAGAGCTTCGAGAGGACGGG	780
Db	4649	CGACCAAGATGTTAGATTTTTCAGCAAGTTACAGCGCTCTAAGAGCTTTTGAGAGGACGGG	4708
Qy	781	AGGAATCTGACGCTCCCACGGTGACGAAGTCTCTGAGTTACATCGCGCTGGCATCTTCG	840
Db	4709	AGGAATCAGCGCGCTCCACACAGTCAGAAAACTCTTTGAATTCCTCTGCGGCTGGACTCTTCT	4767
Qy	841	CGAAGACTCGGCTCGTCAGTACTCTTGATTTGGCTAGAGGAGGACTCCAGTGGCAGTGC	900
Db	4768	TCGGAAGACTTGGTTGTTCAGTTCCTCTGATTTGGCTAGAGGAGGACTCCAGGGGCAGTGC	4827
Qy	901	TCGTCTCTCAGA-CTGGAAGTTAAACTCAAGGCCCAGGAAAGAGCTTTTAGACTCTGCAAGC	959
Db	4828	TGCTCTCTCAGACCTTGAATTTGAATTCAGGCCCCAGGAAAGAGCTTTAGACTCTACAGC	4887
Qy	960	CCCACTTTCCTCTGAGACAGCAGCTGTGGGAGTGGAGTTGCTGCCAGTAGCCTTCAA	1019
Db	4888	CACACTTTCCTCTTGGAGCAGCAGCTGTGGGAGTGGAGTTGCTGCCCCGGTAGTCTTCAA	4947
Qy	1020	GCTGCTCTAGTCTCCCAACCGAGAATTTGACTCTTCTCATCTCTGGGCTCTGAGGCTTCAG	1079
Db	4948	ACCGGCTTAGTCTCCCAACCGAGAATTTGGCTCTTCTCATCTCTGGGCTCTGAGTACTCAG	5007
Qy	1080	AGCTTAGGTAAATTTCAAGGTAGTTTCCTATCTCTCTGAAACCTTCTCTACCTGGACTACCTT	1139
Db	5008	AGCTTGGGCAATTTCAAGGTAGTTTTCCTATCTCTCTGAAACCTTCTCTACCTGGACTGCGTT	5067
Qy	1140	CCCCAGTTAGGCTCGCGCTGTACAGACAGCGCTGGAGGTGGCCAGTTTGTGGGTTCCGA	1199
Db	5068	CCCCAGTTTAGAGCTGCGCTGTACAGAACAGCGTTCGAGGTGGCCAACTTGTGGATTTCCGA	5127
Qy	1200	ACACTAAACCCAGAGAGCTGTATCTTTCTGAAAGATGGTTGTCAACCTCAGCCGTTGCGG	1259
Db	5128	ACACTATCCCCAGAGAGCTGTACTCTTTCTGAAGATGTTGTCAACCTCAGCCGCTGTGT	5187
Qy	1260	GCAGAGATGTCGGCAAACCGCCTCGAGAAGGTGTCTCGCTCTCTCTACAGAAGGCTGCGG	1319
Db	5188	GCTCAGATGTCCGCAACTGCGCTGCGGAGGTGTCCGCCCTCTCTACAGAAGGCTGCGCA	5247
Qy	1320	GAATTCACCAACCTTCGTTATGAAACGGCTAGAAATTCCTTCAGGCTTAACAAAGGCGAAG	1379

BD160691  
LOCUS BD160691 2942 bp DNA linear PAT 17-JAN-2003  
DEFINITION Primer for synthesizing full-length cDNA and use thereof.  
ACCESSION BD160691  
VERSION BD160691.1 GI:27866449  
KEYWORDS JP 2002191363-A/15534.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Bukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 2942)  
Ota.T. Isogai, T., Nishikawa, T., Hayaishi, K., Saito, K., Yamamoto, J., Ishii, S., Sugiyama, T., Wakamatsu, A., Nagai, K. and Oeuki, T.  
TITLE Primer for synthesizing full-length cDNA and use thereof  
JOURNAL Patent: JP 2002191363-A 15534 09-JUL-2002;  
HELIIX RESEARCH INSTITUTE  
OS Homo sapiens (human)  
PN JP 2002191363-A/15534  
PD 09-JUL-2002  
PF 28-JUL-2000 JP 2000280990  
PI TOSHIO OTA, TAKAO ISOGAI, TETSUO NISHIKAWA, KOJI HAYASHI, KAORU  
PI SAITO,  
PI JUNICHI YAMAMOTO, SHIZUKO ISHII, TOMOYASU SUGIYAMA, AI WAKAMATSU,  
PI KEIICHI NAGAI, TETSUJI OTSUKI  
PC  
C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/ PC  
10,  
PC C12P21/02, C12P1/68, C12P21/08, G06F17/30, C12N15/00, C12N5/00 CC  
Primer for synthesizing full-length cDNA and use thereof FH Key  
Location/Qualifiers  
FT CDS Location/Qualifiers  
1. .2942  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
ORIGIN  
Query Match 24.2%; Score 1320.8; DB 6; Length 2942;  
Best Local Similarity 72.5%; Pred. No. 0;  
Matches 1962; Conservative 0; Mismatches 662; Indels 84; Gaps 16;  
Qy 151 CCGTCGCGCGCGCGGAGGAGGCTTCTCTATGTGGAGCGATCTCACAGCGCTAGG 210  
Db 96 CCATCGCGGAGCGGTGGGGGACTCTCTATCCATGTGTGTTGAAGCGTCGAGCGGACTAGG 155  
Qy 211 ACGTCTCTCTCTAGCGGAGTGGACTACCGCGTGGCGCGCTGCGCGCGGCGCTC 270  
Db 156 GAACCTCTCTCTCC-CGCCAGGATGGAAGTCGATCAGTCGCGCGCTATTTGGCGGGCTGT 214  
Qy 271 TGGGCGGTCTCGGTGCAGCACCTGTTGGGAGCGCGCGCTCTCTGGGCGCTCTCTGCGG 330  
Db 215 TCTTCCCTGTGTTCTGCGCGCGCTGCGCGATTCGCTGCGCTCTGTGGCTTTCTGCTGG 274  
Qy 331 CGCGGGAATCGGATCGGATGCCATCCGCTGGCTGGCGAAGGCGGAGACTGTGTAGACC 390  
Db 275 CTCGAAGATCGGCTGGAGCGAGCGACCGCTGGGCAAGGCGGAGACTCTGTAGGCT 334  
Qy 391 TCGGATCCAGCTCGGCTGACCGCGCTGAGCTCTCTCTCTCTCTCTCTCTGAGAGCGGCC 450  
Db 335 TCCTCCGAATCCCGTCGACCTCCAGCGCGCTGAGCGCGCGCGCGCTTACCTGAGAGACTGT 394  
Qy 451 AAG-GAAGGAGATGGAGACAGGAAACGACAGGCGCGGAGCGCGCTGGCGCTCGGCTG 509  
Db 395 AAGAAAGGAGATGGAGCGCGGAGCGGAGCGGATCGCGGAAACGCGCTTGGCGCTCGGCG 454  
Qy 510 GGCTCTCGGTTCCGCTGCC---TTCTCTCGGAGATCGGACCGCTCTCTCTCGGAGTTC 566  
Db 455 GGCTCTCGGTTCTGCGCACCTTTTCTCTCGGAGATCGCAAGCGGCTCTCTTAAGTTC 514  
Qy 567 CCGCGCGCTTCTCTCGCAAAATCCCGGGAACCTC-----GCTCTGCC 611  
Db 515 CCGACGCGCTCTTGGCGCGGAAATCTCGGGAACCCCACTGCTTCTCTGCGCGAGGCC 574

Qy 612 GAGCGTCGAGACAGGTACTGGACCAAAATGCTTTCTCAGCTCCTTGCCTCTGCTCCCTAGC 671  
Db 575 GAGACTCGGGTCAGTTACTTACGACGAAATGCTCTCCAGCTCCTTGGCGCGCTCCCGGA 634  
Qy 672 CTATTCCAGAAAGCTGCTGCTTTTCGAGCCAGCTTTTCGGGGGCGCTGATTCCTACAGATGG 731  
Db 635 TTGCTTCAGAAAGTGTCTAATTTTCGAGCCAACTTTTCGGTGGAAATGTTTCCGACAGATGG 694  
Qy 732 CTAGATTTTTCGCCCAAGTTTACAGCGCCCTGAGAGCTTTTCGAGAGGACGGGAGGAATCTGAC 791  
Db 695 CTAGATTTTTCGCTGAGTCTACAGCGCCCTGAGAGCCCTGAAGGAGCGGAGAAACAGGCC 754  
Qy 792 GCTCCCAAGGTGAGAGTCTCTGAGTTTACCTGGGCTGGAGCTCTTTCGGGAAAGACTCGC 851  
Db 755 GCGCCCAAGCGAGAAATCTTTGAGTT-CGCTGACGCTCGACTCTCTC---AGACCCCTC 810  
Qy 852 GTCGTCTAGTACTCTTCGATTTGGCTAGAGGAGGAGCTCCAGTGGCAGTGTCTGCTCTCAGA 911  
Db 811 GGTACACAGTCCCTCTGATTTGGCTAGAGGAGGAGTCCACTGGCAATACTTCGCCCCCGA 870  
Qy 912 -CTGGAAGTTAAACTCAAGGCCCGAGAAAGAGCTTTAGACTCTGACAGCGCCCACTTTTCT 970  
Db 871 CCTAAATTTGAGCTTTAAGSCCAAGGAGTGTCTTTGGACCTTCGAGCAGCAGGCTTTCT 930  
Qy 971 CTTGAGCAGCAGCTGTGGGAGGTGGAGTTGCTGCCAGTAGCTTCAAGCTGCTCTAGT 1030  
Db 931 CTTAGAGCAGCAGCTGTGGGAGGTGGAGCTGTTGCCAGTAGCTTTCAATTCCTGCTGTA 990  
Qy 1031 CTCCACCGAGAACTTACCTCTCTCTGCGGCTCTGAGGCTTCAGAGCTTCAGAGCTTAGGTAA 1090  
Db 991 CTCTAACCGGAACTTTGGCTCTTGGCCCTCTGAGGCTCTAAACATTTCAACGATAGACGA 1050  
Qy 1091 TTTCAAGGTAGTTTCTCTATCTCCTGAACCTTCTACTGAGCTTACCTTCCCGAGTTAGG 1150  
Db 1051 TTTGAGTGTGATCTCTATTTGCTGAACCTTCTACTGAGCTGCTTTCCAGGCTAGA 1110  
Qy 1151 GCTGCGCTGTGAGCAGCAGCTGGAGTGCGGAGTGTGTTGGTTCGGAACACTTAACCCC 1210  
Db 1111 AGTCAGCTATCAGAAACAGTGTGAAATAGCAGGAGTAGTCGGCTTCCAGACACTTAACCCC 1170  
Qy 1211 AGAGGCTGTATCTTCTGAAAGTGTGCTGCTCAGCTCAGCTGCTGGCGGCGAGAGTGTG 1270  
Db 1171 AGAGAGCAGCTGCTGAGAGGAGCAATTCATCTCCCGCTGAGTGTGAGTGTGAGTGTG 1230  
Qy 1271 GGCAACCGCTGGAGAGGCTGCTGCTCTCTCTACAGAGGCTGCTGCGGAAATTCACCA 1330  
Db 1231 TCCGCGCTGTGCGAGGAGTGTCCACTCTTTCTACGAGGAGCTTACAGAAATTCACCA 1290  
Qy 1331 CTTTCTGATGAAACCGCTAGAAATTCCT---CGAGGCTTAAACAAAGGCGAAGTGTACCCAC 1387  
Db 1291 TCTTCGCTGAAACCGCTGGAATTTCTTCAACAGGCTAGCAAGGGGCGAGATTTACCCAC 1350  
Qy 1388 CCGTGCACAGATTAATGCGCTATCATAGCTGGAGGAGGACATTAACCTTCTCCGGATGGA 1447  
Db 1351 CCCTGACAGGATTAATGCTTACCAAGCTGGAGGAGGAAACAGAGCTTCTCCGGATGGA 1410  
Qy 1448 CCACAACTATTCACAGATAAACCCAGCACAGCGGCTGCTCCCTGCTGACAGACAGGCC- 1504  
Db 1411 TCCAAACACTGACAGATTAACCCACACAGATTTGTTCTGCTGCTGAGACATTCCTGG 1470  
Qy 1505 -----GGAGCCCACTGAGAAAAAACCAAGAAATTTGGTGTGATTCAGAAAGTTTCAC---- 1551  
Db 1471 AAACACCCAGGAATCCACTGAAGAAAAAATAGAAATTAATACTACAGAGGTTCCACTTGC 1530  
Qy 1552 -----AGAGCCCCCAGGAGGAGGAGCTGTTTGTGTAATTAACCTGCGGAAAGA 1600  
Db 1531 TTTGGAAGAGAGAGGCGCTTCTGAGGCTGTCCATCTAGTGAGATACCTATGAGAAAGGA 1590  
Qy 1601 ATGTGAAGAGGACACACTAATCACTGACCTCTCAGATAGAGGAGAGAGGCTTCTCTGT 1660  
Db 1591 GCCTGGAGAGGCGGAATAAGTGTGATTGATTACTCATACTAGAGGAGGAGGCTTCCAT 1650

```
QY 1661 TTCTACACGACGAGTTGTAGCAACAACTGATAGATTATATTTTGGGAGGCGCCCCCAG 1720
Db 1661 TTCTGCCAGACGAGCTTGTAGTAAACAACCTGATAGATTATATTTTGGGAGGTCATCCAG 1710
QY 1721 TGACTTGGAGCGAGCTCTGATTCTGAAAGTCAGGATTGGGCGGAGGAACCTGAGGAGCA 1780
Db 1711 TGACTTGGAAACAAGTCTCGATCCAGAAGGTGAGGATTGGGATGAGGAAGCTGAGGATGA 1770
QY 1781 TGGCTTTGATAGCGATGGCTCCCTCTCTGAATCAGACCTGGAACAGGACTCGGAAGGCT 1840
Db 1771 TGGTTTGTAGTAGTAGTCACTGTCTGAGCTCAGACCTTGACACAGACCTCGAAGGCT 1830
QY 1841 TCACCTTTGGAACTCTTTTCCAGAGTGTAGATCTTTTACAAACCCCAAACTTTTACAGCCAC 1900
Db 1831 TCACCTTTGGAACTCTTTCTGAGTGTAGATCTTTTATAATCCCAAGAACTTTTACAGCAAC 1890
QY 1901 GATTCAGAGCGTGCAGAAATTTGCCCCAGAGACCCATCAGATTCAGGAGCATCTGTGTC 1960
Db 1891 AATTCAGACTGCTGCCAGAAATTTGTTCTTGAAGAGCCTTCTGATTCAGAGNAGGATTTGTC 1950
QY 1961 TGGCAGCTGTGTGTAGGCA---GCTGTCAGGAGGACCCCTTCCGAGAGACCCCGACCA 2017
Db 1951 TGGCAAGTCTGATCTAGAGAAATCTCTCCAGTCTGGAAGCCTTCTGAGACCCCTGAGCA 2010
QY 2018 TAGTTCCGGGAGGAGATGACTGCGGAACCGAGTGCAGATGAAGCAGAGAACTTTAAAT 2077
Db 2011 TAGTTCTGGGAGGAGATGACTGCGGAATCTAGTGCAGATGAAGCAGAGATCTCAAACT 2070
QY 2078 GTGGAACCTTTCTGTCATCTGAGGACCCCTACAACTTTTAAATTTTAAAGGCTCCTTT 2137
Db 2071 GTGGAACCTCATTTCTGTAATCTGATGACCCCTTACAACTTTTAAATTTTAAAGGCTCCTTT 2130
QY 2138 TCAACCGTTCAGGAGAAATTTGGAAGGCGCTCAGGACTCAAGGCTCTTCTGAGGTCCAC 2197
Db 2131 TCAAAATCAGGAGAAATTTGAAGGCTGTGTCATCAAGACCCCATCTGAGTCCAT 2190
QY 2198 AGTGCCCTTCTCTGGCCATCATACCTTCTTGTGAAGCCCAAGCTGTTTAGAGAGCA 2257
Db 2191 TGTGGCCATTTCTGAGTGTCAACCTTACTTCTTGTGAAGTGCAGCTGTTGGGAGGCA 2250
QY 2258 AGAAGATAATTTGTCAGGCTGTGGCTGGGTGAGGCTTCTTCTGAGGAGAAATACACCA 2317
Db 2251 AGAAAGTGAATTTCCAGACTCGGTAACGCTGAGCTTCTTCTGAGGAGAAACACACA 2310
QY 2318 TATCAAGAGAAAAAGTGAACCTTCTCGAAGAGTCTACTGAGTATTATATAAGTGCTGA 2377
Db 2311 TGTCAAAAGAAAAAGGTAACTTCTTGAAGAGTACTGAGTATTATATAAGTGCTGA 2370
QY 2378 TGAGGATCCGAAGGACCATGGGAAGAAATTTGCAAGGATGGATGCAGGTTCCAGAAACG 2437
Db 2371 TGAGGATCCGAAGGACCATGGGAAGAAATTTGCAAGGATGGATGCAGGTTCCAGAAACG 2430
QY 2438 AATTCAGAAACAGAAAGTTGCCATTTGGCTACTGCTTGGCCCTTTGAGCACAGAAAAAT 2497
Db 2431 AATTCAGAAACAGAAAGTGTATTGGATATTGCTTGAATTTTGAACACAGAGAAAGAT 2490
QY 2498 GTTTAATAGACT---GAGGATCGAGTCAAAAGGACTTACTGTGTGACAGCAATGTTAAGA 2553
Db 2491 GTTTAATAGACTCCAGGGAACATGCTTCAAAAGGACTTAAATGTTCTCAAGCAATGTTAGT 2550
QY 2554 AGTGA-ACAGCCTGCAACCGGTGCCACTCTGTCTCTTTACTTTGAGA-GTTTTCCCTTAAA 2611
Db 2551 TGGCAGCTGTAGTCTAGCTAGCATACACTACCTCTTACCTGAGAGGTGCTTTTAAAA 2610
QY 2612 ACAACACTGGGAGCTGCTTGGACATGTTTTTAAAGAAACAACATTGTATCTAGAGATG 2671
Db 2611 ACAAACTTGGCAGCTGCTTGTGACATTTTTTTTTTTTATAGGAAATGTAACTTTGGATCT 2670
QY 2672 CAGTTTGAATTTTTTTGGTAAATGTCTCATTTAGAA-----ACACCAACT 2717
Db 2671 AGTTTAAATTTTTTTTTTTTGGACATATCCCACTCAGAAACAATTCAGGTTTGAAGCCGCC 2730
QY 2718 CCGATAATGAAGAAAT-----CTCTTATCTGTAATCTCTCTCTTTTCTTATTTAGTGGAT 2771
```

```
Db 2731 CTGATAATGAAGGATGAAGTAGTGTGATTTCTAATCTCCTCTTTTGAATTAGTGGAT 2790
QY 2772 GTGGGTTT 2779
Db 2791 GTGCTTTT 2798

RESULT 10
AX884059 2942 bp DNA linear PAT 17-DEC-2003
LOCUS AX884059 Sequence 18964 from Patent EP1074617.
DEFINITION AX884059
ACCESSION AX884059
VERSION AX884059.1 GI:40038960
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Mukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Ota,T., Isogai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J.,
Ishii,S., Sugiyama,T., Wakamatsu,A., Nagai,K. and Otsuki,T.
TITLE Primers for synthesising full-length cDNA and their use
JOURNAL Patent: EP 1074617-A 18964 07-FEB-2001;
RESEARCH Association for Biotechnology (JP)
FEATURES
Location/Qualifiers
1..2942
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
407..2548
/notes="unnamed protein product"
/codon_start=1
/protein_id="CA892217.1"
/db_xref="GI:40038961"
/translation="MEFGTGSRRKRLGPRAGFRFWPFFRRRSQAASSKFPPLGPEN
SGNPTLLSSAQPETRVSYMTKLISQLLAPLGLQKVLVWSLFGMFPFRWLDFAGV
YSALRALGKREKPAFTAQKLSLQSDSDSPVTPSLDWEIHWQSPPLDKUEL
KAGSALDPAQAFLLEQLWGVLLPSSLSRLYSNRELSSPSGPLNATQRIDDFSV
VSYLNPVLDLCPFLVSVYONSDGNSVVGFTLTPSSCLREDHCHGPOLSAELTP
ASWQGPCPLSTEGLEPIHLRMKRLRFLOQASKQDLPTPDQDNGVHSLREHSLLRM
DPKXDRNDPQFVPAAGDIPGNTQESTEEKIELLTTEVPLALEEESPSECPSEIPLM
EKPGEGRISVDVSYLEGLDPLISARPACSNKLIDYILGGASDLETSSDPFEGDWE
EADDGFDSDSLDLEQDPEGLHWNFSFVYNPQNFTATOTAAARIVPPEES
DSKFLSGKSDLENSQSLPETPSSGDEDDWESSADEASLKLWNSFCNDSYQ
NPLNFKAPFOTSGENEGKCRDSTPSESIVAISECHTLLSKVQLLGSQSECPDSYQ
RDVLSGGRHTVKKRKYVLEVEVTVYISGEDRDKGPEEFARDGCRFQRIQETEDA
IGCLTFEHRHRMFRNLQGTCTFKGLNVLKQC"
```

## ORIGIN

```
Query Match 24.2%; Score 1320.8; DB 6; Length 2942;
Best Local Similarity 72.5%; Pred. No. 0;
Matches 1962; Conservative 0; Mismatches 662; Indels 84; Gaps 16;

QY 151 CCCTGCGCGCGCGGAGGAGGTCTTCTTATGTTGGAGCGATCTCACACGCGCTAGG 210
Db 96 CCATGCGCGAGGCGGTGGGGGAGCTCTATCCATGGTGTGAAGCGTCGAGCCGACTAGG 155

QY 211 ACGTCTCTTCCCTAGCCGGATGGACCTAAACCGGTCGCCACCGCTTGGCGGGCCTC 270
Db 156 GAACCTCTTCCC-CGCCAGGATGGAAGTCGACATCAGTCGCGCTTATTCGCGGGGCTGT 214

QY 271 TGGGCGCTCCGGTGCAGACTCTGTTGGGAAGCGCGCGCTCTCTGGGCGCTCTCTGCGCG 330
Db 215 TCTTCCCTGTGTTCTGCGCGCGCGCTGCGCATTCGCTGCGCTCTGTGGCTTTTCTGCTGG 274

QY 331 CGCGGGAATCGGACTGCAGTACCCACTCCCTGCTGGCTGGGCAAGCGGAGAGACTGTGTAGACC 390
Db 275 CTCNAGATCGGCTGCGAGCAGCGACGCCACCGCTGGGCAAGGCCGAGACTCTGTAGGCT 334

QY 391 TCGGATCCAGCCTGCGCTGAGCGCGCTGAGCTCTGTCTCTCTCTCTGCTCAGAGAGCCGCC 450
Db 335 TCCTTCGAAATCCCGCTCGACTCTCAGCGCGCTGAGCGCGCGCGCTTACCTGAGAGACTGTC 394
```





Db 2551 TGGCAGCCTGTAGTCTCTAGTACATACATACCTCTTACCTGAGAGGTGTCTTTTAA 2610  
Qy 2612 ACAACACTGGCAGCTGCTCTGGACATGTTTTTAAAGAAACAACCTGTACTAGAGATG 2671  
Db 2611 ACAATCTTGGCAGCTGCTCTGGACATGTTTTTAAAGAAACAACCTGTACTAGAGATG 2670  
Qy 2672 CAGTTTGATTAATTTTTGGGTAAATGTGTCTCATTTAGAA-----ACACCAACT 2717  
Db 2671 AGTTTAATTTTTTTTTTGCAACATATCCCACTCAGAAACATTCAGGTTTGAAGCCAGCC 2730  
Qy 2718 CGATAATGAAGAAAT-----CTCTATCTGTATATCTCTCTCTTTTCCCTATTAGTTGAT 2771  
Db 2731 CTGATAATGAAGGATGAAGTGTGTGATTTCTTAATCTCTCCCTTTTTCATTTAGTTGAT 2790  
Qy 2772 GTGGGTTT 2779  
Db 2791 GTGCTTTT 2798

RESULT 11  
AK027650  
LOCUS Homo sapiens cDNA FLJ14744 fis, clone NT2RP3002770. linear PRI 30-JAN-2004  
ACCESSION AK027650  
VERSION AK027650.1 GI:14042484  
KEYWORDS oligo capping; fis (full insert sequence).  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1  
Ota, T., Suzuki, Y., Nishikawa, T., Otsuki, T., Sugiyama, T., Irie, R.,  
Wakamatsu, A., Hayashi, K., Sato, H., Nagai, K., Kimura, K., Makita, H.,  
Sekine, M., Ohayashi, M., Nishi, T., Shibahara, T., Tanaka, T.,  
Ishii, S., Yamamoto, Y., Saito, K., Kawai, Y., Isono, Y., Nakamura, Y.,  
Nagahari, K., Murakami, K., Yasuda, T., Iwayanagi, T., Wagatsuma, M.,  
Shiratori, A., Sudo, H., Hosiiri, T., Kaku, Y., Kodaira, H., Kondo, H.,  
Sugawara, M., Takahashi, M., Kanda, K., Yokoi, T., Furuya, T.,  
Kikkawa, E., Omura, Y., Abe, K., Kamihara, K., Katsuta, N., Sato, K.,  
Tanikawa, M., Yanagaki, M., Ninomiya, K., Ishibashi, T., Yamashita, H.,  
Murakawa, K., Fujimori, K., Tanai, H., Kimata, M., Watanabe, M.,  
Hirooka, S., Chiba, Y., Ishida, S., Ono, Y., Takiguchi, S., Watanabe, S.,  
Yosida, M., Hotta, Y., Kusano, J., Kanehori, K., Takahashi, Fujii, A.,  
Hara, H., Tanase, T., Nomura, Y., Togiyama, S., Komai, F., Hara, R.,  
Takeuchi, K., Arita, M., Imose, N., Musashino, K., Yuuki, H., Oshima, A.,  
Sasaki, N., Aotsuka, S., Yoshikawa, Y., Matsunawa, H., Ichihara, T.,  
Shohata, N., Sano, S., Mori, Y., Momiya, H., Satoh, N., Takami, S.,  
Terashima, Y., Suzuki, O., Nakagawa, S., Senoh, A., Mizoguchi, H.,  
Goto, Y., Shimizu, F., Wakebe, H., Hishigaki, H., Watanabe, T.,  
Sugiyama, A., Takemoto, M., Kawakami, B., Yamazaki, M., Watanabe, K.,  
Kumagai, A., Itakura, S., Fukuzumi, Y., Fujimori, Y., Komiyama, M.,  
Tashiro, H., Tanigami, A., Fujiwara, T., Ono, T., Yanada, K., Fujii, Y.,  
Ozaki, K., Hiro, M., Ohmori, Y., Kawabata, A., Hiki, T., Kobatake, N.,  
Inagaki, H., Ikem, Y., Okamoto, S., Okitani, R., Kawakami, T.,  
Noguchi, S., Itoh, T., Shigeta, K., Senba, T., Matsumura, K.,  
Nakajima, Y., Mizuno, T., Morinaga, M., Sasaki, M., Togashi, T.,  
Oyama, M., Hata, H., Watanabe, M., Komatsu, T., Mizushima-Sugano, J.,  
Satoh, T., Shirai, Y., Takahashi, Y., Nakagawa, K., Okumura, K.,  
Negase, T., Nomura, N., Kikuchi, H., Masuho, Y., Yamashita, R.,  
Nakai, K., Yada, T., Nakamura, Y., Ohara, O., Isogai, T. and Sugano, S.  
Complete sequencing and characterization of 21,243 full-length  
human cDNAs  
Nat. Genet. 36 (1), 40-45 (2004)

2  
Isogai, T., Ota, T., Hayashi, K., Sugiyama, T., Otsuki, T., Suzuki, Y.,  
Nishikawa, T., Nagai, K., Sugano, S., Takahashi-Fujii, A., Hara, H.,  
Tanase, T., Nomura, Y., Togiyama, S., Komai, F., Hara, R., Takeuchi, K.,  
Arita, M., Nabekura, T., Ishii, S., Kawai, Y., Saito, K., Yamamoto, J.,  
Wakamatsu, A., Nakamura, Y., Nagahari, K., Masuho, Y. and Oshima, A.  
NEDO human cDNA sequencing project  
Unpublished

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

COMMENT

FEATURES  
source

CDS

ORIGIN

Query Match 24.2% Score 1320.8; DB 9; Length 2942;  
Best Local Similarity 72.5% Pred. No. 0;  
Matches 1962; Conservative 0; Mismatches 562; Indels 84; Gaps 15;  
Qy 151 CCCTGCGCGCGCGGAGGAGGTCTTCTTCTATGTTGGAGCGATCTCACACGCGCTAGG 210  
Db 96 CCATGCGCGGCGCGTGGGCGGACTCTCTATCCATGGTGTGAAAGCGTGGAGCGGCTAGG 155  
Qy 211 ACCTCTCTCTCTAGCGCGGATGGACCTAACCGCGTGGCGCGCTTGGCGGGGCTC 270  
Db 156 GAACCT 214  
Qy 271 TGGGCGCGTGGGCGGAGCGAGCTCTGTTGGGAAAGCGCGCTCTCTGGGCGCTCTCTGCGG 330  
Db 215 TCTTCT 274  
Qy 331 CGCGGGAATCGGATCGAGTACCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 390  
Db 275 CTCGAGATCGGCGCTGGGCGGAGCGGACCGCTGGGCGGAGCGGCGGAGCTCTGTAGGCT 334  
Qy 391 TCGGATTCAGCTCGGCTGAGCGCGCTGAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 450  
Db 335 TCTCTCGAATCT 394  
Qy 451 AAG-GAAAGGAGATGGAGACAGGAAACGACAGCGCGCGCGGAGCGGCTCTGCGCTCGGCTG 509  
Db 395 AAGAAAGGAGATGGAGCGGCGGAGCGGCGGAGCGGCTCTGCGCTCTGCGCTCTGCGGCG 454

3 (bases 1 to 2942)  
Isogai, T. and Otsuki, T.  
Direct Submission  
Submitted (10-MAY-2001) Takao Isogai, Helix Research Institute,  
Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan  
(E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)  
NEDO human cDNA sequencing project supported by Ministry of  
Economy, Trade and Industry of Japan; cDNA full insert sequencing:  
Research Association for Biotechnology; cDNA library construction,  
5'- & 3'-end one pass sequencing and clone selection; Helix  
Research Institute (supported by Japan Key Technology Center etc.)  
and Department of Virology, Institute of Medical Science,  
University of Tokyo.  
Location/Qualifiers  
1. .2942  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="NT2RP3002770"  
/cell\_line="NT2"  
/cell\_type="teratocarcinoma"  
/clone\_lib="NT2RP3"  
/notes="Cloning vector: pME18SFL3-mRNA from NT2 neuronal  
precursor cells after 2-weeks retinoic acid (RA)  
induction."  
407..2548  
/notes="unnamed protein product"  
/codon\_start=1  
/protein\_id="BAB55266.1"  
/db\_xref="GI:14042485"  
/translation="MEFGTGGSRKRLGPRAGFRFWPFFPRRQAGSKKPTPLGPEN  
SGNPTLSSAQPTRYWTKLSQLLPGLLQKVLWLSQFGLGMPFRWLDFAV  
YSALRAKGRKPAAPTAQKSLSSLDSDSPSVTSPDLWEIGIHQWYSPDLKLEL  
KAGSALDPAQAFLLEQLGWVLLPSSLSQSLRYNRELSSGSPGLNQRIDDFSV  
VSYLLNPSYLDPCPRLEVSQNSGSENVVGFOTLTPESCLREDHCHPCLARLIP  
AQWCGPPLSTEGPIHHLMKRLEFLQASKGQDLPTPDQNGHSLSEESLLAM  
DPKQNDNPTQFVPAAGDIPGNTTEETKEILLTTEVPALAEESPSGCGPSEIIPM  
EAEEDGFSVVDYSYLEDPLTSARPACSNKLDYILGGASDLSTTSAAIIVPEEPS  
DSKLLSGSDLENSQSGLPTPEHSSGEEDWESSADEASLKLWNSFCNSDDPY  
NPLNFAPQVTCSENEKGRDSTPSEIIVAISECHLLSKVQLLGSQSECPDSVQ  
RDVLSGRTHVRKKVTFLEEVTEYIISGDEDRKGPWEFEFARDGCRFQRIOETEDA  
IGYLTFFEHRRFMNRLQGTCTCFKGLNVLKQC"

Qy 510 GGCTCTGGTTCCGGCTGCCCC---TTCTCTCGCGGATCGCAGCGCTGCTCTTCGAGATTC 566  
Db 455 GGGTTCCGGTTCTGGCCACCTTTTTCCTTCGCGGATCGCAAGCAGGCTCTTCTAAGTTC 514  
Qy 567 CCGCGGCTTCTCTCGACAAATCCCGGAATCCG-----GCTCTGCC 611  
Db 515 CCGACGCTCTTGGCCCGGAATCTCGGGAACCCCACTGCTTTCTGCGCCAGCCC 574  
Qy 612 GAGCGTCGACACGAGTACTTGACCAAAATTTCTCTCAGCTCTTTCGCTGCTCCCTAGC 671  
Db 575 GAGACTCGGTCAGTTACTTGACGAAATCTGCTCCCGAGCTCTTGGCGCTCCCGGA 634  
Qy 672 CTATTCAGAGCTGCTGCTTTTGAGCCAGCTTTCCGGGGCCCTGATTTCTTACAGATGG 731  
Db 635 TTGCTTCAGAGGTGCTAAATTTTGAGCCAAATTTTTCGGTGGAAATGTTTTCGACCAATGG 694  
Qy 732 CTAGATTTTCCCGCAAGTTACAGCGCCCTCAGAGCTTCGAGAGACGGGAGGAATCTGAC 791  
Db 695 CTAGATTTTCTGAGGTCTACAGCGCCCTGAGAGCCCTGAAGGACCGGGAACCAAGCC 754  
Qy 792 GCTCCCAACGGTGCAGAAATCTCTGAGTTACACTCGCGCTGGAATCTTTCGCGAAGACTCGC 851  
Db 755 GCCCCACACGCGAGAAATCTTTGAGTT- GCTGCACTCGACTCTCTC---AGACCCCTC 810  
Qy 852 GTCGTCAGTACTCTTGCAATTTGGCTAGAGAGGAGCTCCAGTGGCAGTGTCTGCTCTCAGA 911  
Db 811 GGTCAACAGTCCCCTTGATTTGGCTAGAGGAAGGATCCACTGGCAATACTCGCCCCAGA 870  
Qy 912 -CTGGAAGTTAACTCAAGCCCAAGGAAGCTTTAGACTCTGCAAGCCCACTTTCCT 970  
Db 871 CCTAAAAATTGAGCTTTAAGGCCAAGGGAAGTCTTTTGACCCCTGCAAGCAAGGCTTTTCT 930  
Qy 971 CTTGAGCAGCAGCTGTGGGGAGTGGAGTTGCTGCCAGTAGCCCTCAAGCTGCTTAGT 1030  
Db 931 CTTAGACAGCAGCTGTGGGGAGTGGAGCTGTTGCCAGTAGCTTCAATCCCGTCTGTA 990  
Qy 1031 CTCCCAACGAGAACTTGACTCTTCATCTCTGCGCCCTCTGAGCGTTTCAGAGCTTAGGTAA 1090  
Db 991 CTCTAACCGGAACTTGGCTCTTGGCCCTCTGAGGCTCTAAACATTTCAAGGCATAGACGA 1050  
Qy 1091 TTTCAAGTATTTCTCTATCTCTGAACCTTCTCTACAGAGGCTGCGGAAATCCACCA 1150  
Db 1051 TTTCAAGTATTTCTCTATCTCTGAACCTTCTCTACAGAGGCTGCGGAAATCCACCA 1110  
Qy 1151 GCTGCGCTGTGAGAGCAGCTGAGGTGGCCAGTTTGTGGTTTCCGAACTAACCCCC 1210  
Db 1111 AGTCAGCTATCAGAAAGTATGGAATAGGGAGTGTGGCTTCCAGACATTAACCCC 1170  
Qy 1211 AGAGAGCTGTATCTTTCTGAAGATGTTGTCAACCTCAGCGGTTGCGGGCAGAGATGTC 1270  
Db 1171 AGAGAGCAGCTGCTGAGAGAGGACCAATTTGTATCTCCCGAGCGCTGAGTGCAGAACTCAT 1230  
Qy 1271 GGCNAACCGCTGAGAGAGTGTGCGCTCTCTCTACAGAGGCTGCGGAAATCCACCA 1330  
Db 1231 TCCGGCTCTGTTGGCAGGATGTCCACCTCTTTTACGGAAGGCTTACAGAAATTCACCA 1290  
Qy 1331 CCTTCGTATGAAACGGCTAGAAATTCCT---CCAGGCTAAACAAAGGCAAGAGTTTACCCAC 1387  
Db 1291 TCTTCGATGAAACGGCTGGAATTCCTTCAAGGCTAGCAAGGGGCAAGATTTACCCAC 1350  
Qy 1388 CCCTGACCAAGATATGCTATATAGCTGGAGAGGAACATAACCTTCTCCGGATGGA 1447  
Db 1351 CCCTGACCAAGATATGCTATACAGCTGAGAGGAGCAACAGACCTTCTCCGGATGGA 1410  
Qy 1448 CCGAACAATGACAGATAACCCAGCAGCGCGGTGCTCCCTGCTCTCAGACAGGCC--- 1504  
Db 1411 TCCAAAACACTGACAGATAAACCCCAACAGATTTGTTCTCTGCTGCTGAGACATTTCTCTGG 1470  
Qy 1505 -----GGAGCCCACTGAGAAACCAAGAAATTTGTTGATTCAGAAAGTTTTCAC--- 1551  
Db 1471 AAACACCCAGGAATCCACTGAGAGAAANATAGAAATTAATTAACAGAGGTTCCACTGCG 1530  
Qy 1552 -----AGAGCCCCCAGGGAAGCAGTCTGTTTGTGAAATTAACCGGTGGAAAAAGA 1600

Db 1531 TTTGGAAGAGAGAGCCCTTTCTAGGGCTGTTCATCTAGTGAGATACCTATATGAAAAAGGA 1590  
Qy 1601 ATGTGAAGAGGAGCAACACATAATGCAACTGACCTCTCAGATAGAGGAGAGAGCCCTTCTGT 1660  
Db 1591 GCCTGAGAGGGCGGAATAAGTGTAGTTACTCATACCTAGAAGGTGACCTTCCCAT 1650  
Qy 1661 TTCTACAGACACAGTTTGTAGCAACAACTGATAGATTAATATTTTGGGAGGCGCCCCAG 1720  
Db 1651 TTCTGCGACAGCAGCTTGTAGTAACAACTGATAGATTAATATTTTGGGAGGTGCATCCAG 1710  
Qy 1721 TGACTTTGGAAGCCAGCTCTGATTTCTGAAAGTGAAGGATGCGGCGAGGAACCTCAGAGACGA 1780  
Db 1711 TGACCTTGGAAACAAGTTCTGATCCAGAGGTGAGGATGCGGATGAGAAAGCTGAGGATGA 1770  
Qy 1781 TGGCTTTGATAGCGATGGCTCCCTGTCTGAAATCAGACGCTGGAAACAGGACTCGGAAGCCCT 1840  
Db 1771 TGGTTTGTATAGTGTAGCTCACTGTCTGAGCTCAGACTGGAACAGACCTTGAAGGGCT 1830  
Qy 1841 TCACCTTTGGAATCTTTTTCACAGTGTAGATCTTTACAAACCCCAAAACCTTTACAGCCAC 1900  
Db 1831 TCACCTTTGGAATCTTTTTCGAGTGTAGATCTTTAATATCCCCAGAACTTTACAGCAAC 1890  
Qy 1901 GATTTCAGACGGCTGCCAGAAATTCGCCCCAGAGACCCATCAGATTTCAAGGACATCTCTGGTC 1960  
Db 1891 AATTTCAGACTGCTGCCAGAAATTTGTTCTGAAAGAGCCCTTCTGATTCAGAGAAGGATTTGTC 1950  
Qy 1961 TGGCAGCTGTGGTGTAGGGA---GCTGTGAGGAGGACCCCTTCCCGAGACCCCCGACCA 2017  
Db 1951 TGGCAAGTCTGATCTAGAGAAATTCCTCCAGTCTGGAAGCTTCTCTGAGACCCCTGAGCA 2010  
Qy 2018 TAGTTCGGGGAGGAAAGATGACTGGGAAACGAGTGCAGATGAAAGCAGAGAAATCTTAAAT 2077  
Db 2011 TAGTTCGGGGAGGAAAGATGACTGGGAAATCTAGTGCAGATGAAAGCAGAGAGTCTCAAACT 2070  
Qy 2078 GTGGAATCTTTTCTGTCATTTCTGAGGACCCCTCAACCTTTTAAATTTTAAAGCTCCTTT 2137  
Db 2071 GTGGAATCTTTTCTGTAATTTCTGATGACCCCTTAAACCTTTTAAATTTTAAAGCTCCTTT 2130  
Qy 2138 TCAACCGTTCAGGGAAGAAATGGGAAAGCGCTCAGGACTCAAGAGCCCTTCTCTGAGGTCAAC 2197  
Db 2131 TCAAAACATCAGGGGAAATGAGAAAGGCTGTCTGTGACTCAAGACCCCATCTGAGTCCAT 2190  
Qy 2198 AGTGGCTCTCTCTGGCCATCATACCTTATCTTTCTGTAAAGGCCAGCTGTGTAGAGACCA 2257  
Db 2191 TGTGGCCATTTCTGAGTGTACACCTTACTTTCTGTAAGGTGACGCTGTGGGGAGCCA 2250  
Qy 2258 AGAAGATTAATTTCTCAGGCTGTGGGCTGGGTGAGGCTCTGTCTGGAGAAAGATACACCCA 2317  
Db 2251 AGAAGTGAATGTCCAGACTTCGGTA CAGCGTGTGACGTTCTTTCTGGAGGAAGACACACA 2310  
Qy 2318 TATCAAGAGAAAAAAGGTAACCTTCTCTGGAAGAAAGTTACTGTGATATATATAAGTGTGA 2377  
Db 2311 TGTCAAAAGAAAAAAGGTAACCTTCTCTGGAAGAAAGTTACTGTGATATATATAAGTGTGA 2370  
Qy 2378 TGAGGATTCGAAAGGACCAATGGGAAAGAAATTTGCAAGGATGAGATGCAAGTTCCAGAAACG 2437  
Db 2371 TGAGGATTCGAAAGGACCAATGGGAAAGAAATTTGCAAGGATGAGATGCAAGTTCCAGAAACG 2430  
Qy 2438 AATTCAAGAAACAGAAATTTGCCATTTGGCTACTGCTTTGGCTTTGAGCACAGAGAAAT 2497  
Db 2431 AATTCAAGAAACAGAAATTTGGATATTTGCTTGAATTTGAAACACAGAGAAAGAT 2490  
Qy 2498 GTTTAATAGACT---GAGGATCGAGTCAAGAGGACTTACTGTGTGTATACAGCAATTTTAAAGA 2553  
Db 2491 GTTTAATAGACTCCAGGGAACATGCTTCAAGGACTTAAATGTTCTCAAGCAATTTGAGT 2550  
Qy 2554 AGTGA-ACAGCCTGCAACCGTCCCACTCTGTCTCTTACTTGAGA-GTTTCCCTTAAAA 2611  
Db 2551 TGGCAGCTGTAGTCTCTAGCTAGCATACACTACCTCTTACCTGAGAGGTGCTCTTTAAAA 2610  
Qy 2612 ACAAACACTGGCAGCTGCTCTTGGACATGTTTTTAAAGAAACACTTGTATCTAGAGATG 2671



Db 2611 ACAAATCTGGCAGCTGCTCTTGTGACATTTTTTTTTTTAGAGAAATGTAACCTGGATCT 2670  
Qy 2672 CAGTTTGATATATTTTGGTAATGTCTCATAGAA-----ACACCAACT 2717  
Db 2671 AGTTTAAATTTTTTTTTTGGACATATCCCACTGAGAACATTCAGTTTGAGCCAGCC 2730  
Qy 2718 CGATAATGAGAAT-----CTCTATCTGTAATCCTCTCTTTTCCCTATTAGTTGAT 2771  
Db 2731 CTGATAATGAAGATGAAGTACTAGTGTGTAATCTCTAATCCTCCCTTTTTCATTAGTTGAT 2790  
Qy 2772 GTGGGTTTT 2779  
Db 2791 GTGCTTTT 2798

RESULT 12  
BC065280  
LOCUS BC065280 Homo sapiens cDNA clone MGC:74824 IMAGE:6172811, complete cds.  
ACCESSION BC065280  
VERSION BC065280.1 GI:40807094  
KEYWORDS MGC.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 5264)  
AUTHORS Klausner, R.D., Collins, P.S., Wagner, L.H., Grouse, L.H., Derge, J.G.,  
Alschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhac, N.K.,  
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J.J., Heide, F.,  
Diatchenko, L., Maruina, K., Farmer, A.A., Rubin, G.M., Hong, L.,  
Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,  
Scheet, T.E., Brownstein, M.J., Usdin, T.B., Tohiyuki, S.,  
Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,  
Abramson, R.D., Mullany, S.J., Bosak, S.A., McEwan, P.J.,  
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,  
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Huly, S.W.,  
Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,  
Fahey, J., Helton, E., Kettman, M., Madan, A., Rodriguez, S.,  
Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,  
Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,  
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,  
Butterfield, Y.S., Krzyzanski, M.I., Skalska, U., Smalish, D.E.,  
Schnerch, A., Schein, J.E., Jones, S.J., and Marra, M.A.  
Generation and initial analysis of more than 15,000 full-length  
human and mouse cDNA sequences  
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)  
12477932

2 (bases 1 to 5264)  
DIRECTOR MGC Project.  
TITLE Direct Submission  
SUBMITTED (12-JAN-2004) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
Tissue Procurement: ATCC  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILLNL)  
DNA Sequencing by: National Institutes of Health Intramural  
Sequencing Center (NISC),  
Gaithersburg, Maryland.  
Web site: <http://www.nisc.nih.gov/>  
Contact: [nisc.mgc@nih.gov](mailto:nisc.mgc@nih.gov)  
Akhter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B.,  
Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S.,  
Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P.,  
Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Laric, P., Legaspi, R.,  
Maduro, Q.I., Masiello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C.,  
McBowell, J., Pearson, R., Stantripop, S., Thomas, P.J., Touchman, J.W.,

Targeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,  
Young, A., Zhang, L.-H. and Green, E.D.  
Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/ILLNL at: <http://image.llnl.gov>  
Series: IRAK Plate: 140 Row: p Column: 3  
This clone was selected for full length sequencing because it  
passed the following selection criteria: matched mRNA gi: 41282220.

FEATURES  
source

1..5264  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="MGC:74824 IMAGE:6172811"  
/tissue\_type="Eye, retinoblastoma"  
/clone\_lib="NIH\_MGC\_67"  
/lab\_host="DH10B"  
/note="vector: pCMV-SPORT6"  
381..2522  
/codon\_start=1  
/product="Unknown (protein for MGC:74824)"  
/protein\_id="AAH65280.1"  
/db\_xref="GI:40807095"  
/translation="MEFGTGGSRKRLGPRAGFRWPPFRPSQAGSSKFPPLGPEN  
SGNPTLLSSAQPETRVSYTKLISQLLAPLGLQKVLWISQFGMFPTRWLDFAV  
YSALRALKGRKPAAPTAQKSLSLQDSDSPVTSPLDWLEGIHWQYPPDLKLEL  
KAKGALDPAQAFLLEQLWGLVLLPSSLSRLYNSRLGSSPSGSPPLNTQRIDNFSV  
VSVLNPPLSDCFPLRVLSYONGSNVVGFTLTPESLCHDHCHPOPLSAELIP  
ASWQCCPPLSTEGLEPIHLRMRKLFLOAQSGQDLPTPDQNGHLSLEERHLLRM  
DPKCRDNPQFPVAGADIPONTQESTEEKIELLTTEVFLALSESPSECPSEIPM  
EKPEGRISVVDYSLEGLDPLISARPACSNKLIDYLGGASDLSLETSPDEGEDWDE  
EADDGFDSDSLSDLEQDPEGLHLWNSFCSDVPYNQNFATTTQTAARIVPEEPS  
DSKDLKAPFOTSGENKSGSLRPTPEHSGEEDWESSADAEASLLKWSFCNSDDPY  
NPLNFKAPFOTSGENKSGSLRPTPEHSGEEDWESSADAEASLLKWSFCNSDDPY  
RDVLSGGRHTRHKKVTELEVTEYVSGDEDRKGPWEFARDCGRFQKRQIQTEDA  
IGCLIFERRRNFRLQGTGCFKGLNVLKQC"

## ORIGIN

Query Match 24.2%; Score 1320.8; DB 9; Length 5264;  
Best Local Similarity 72.5%; Pred. No. 0;  
Matches 1962; Conservative 0; Mismatches 662; Indels 84; Gaps 16;  
Qy 151 CGGTGCGCGCGCGGAGGAGGAGGTCTTCTATGTGTGAGGAGATCTCACACGGCCTTAGG 210  
Db 70 CCATCGCGGAGCGCGTGGGGGAGCTCTCATCATGTGTGTGAAGCGTCGAGCGGACTAGG 129  
Qy 211 ACGTCTCTTCCCTAGCGCGGATGGACCTAACCGCGTCCGACCGCTTCGCGGGGCGTC 270  
Db 130 GAACCTCTTCCC-CGCCAGATGGAAGTCGATCGATCGCGCGCTATTCGCGGGGTGT 188  
Qy 271 TGGGCGCGTCCGCTGCAGCACTCGTTGCGGAAGCGCGCGCTCTCTGGGCGCTCTCTGCGCG 330  
Db 189 TCTTCCCTGTGTTCTGCGCGCGCTGCGCATTCGCTGCGCTCTGTGGCTTTCTGCTGG 248  
Qy 331 CGCGGGAATCGGACTGCGAGTACCACCTCGGTGGCTGGGCAAGCGGAGAGACTGTGTAGACC 390  
Db 249 CTGGAAGATCGGCTGCGAGCAGCGACCGCCCTGGGCAAGCGGAGACTCTGTAGGCT 308  
Qy 391 TCGGATCCAGCGTGGCTGAGCGCGCTGAGCTCTGTCTCTCTCTGCTCAGAGAGCGGCC 450  
Db 309 TCTTCGGAATCCCGTGCACCTCCAGCGCGCTGAGCGCGCGCGCGCTACCTCAGAGAGCTGC 368  
Qy 451 AAG-GAAAGGAGATGGAGACAGGAAAGCAAGGCGCGCGGAGCGGCGCTCGCGCTG 509  
Db 369 AAGAAAAGGAGATGGAGCGCGGAGCAGGCGGATCGCGGAAACGGCTTGGCCCTCGGGCG 428  
Qy 510 GGCTCTCTGTTCCGGCTGCGCC-----TTCTTTCGGGATCGCACCGCTCTCTTGGAGTTC 566  
Db 429 GGCTTCGGTCTTGGCCACCTTTTTCCTTCGGCGATCGCAAGCAGGCTCTTCTTAAGTTC 488  
Qy 567 CCGCCGCTTCTCTGACAAAAATCCGGGAACTCC-----GCTCTGCC 611  
Db 489 CCGACGCTCTTTCGCGCGGAAACTCGGGAAACCCACACTGCTTCTCTCTGCGCGGCC 548

Qy	612	GAGCGTCGGACCAAGGTACTGGACCAAAATGCTTTCTCAGCTCCTTGGCCCTGCTCCCTPAGC	671
Db	549	GAGACTCGGGTCAGTTACTGGACGAAACTGCTCTCCAGCTCCTTGGCGCGCTCCCGGGA	608
Qy	672	CTATTCCAGAGCTGCTGCTTTGGAGCCAGCTTTCCGGGGGCCGTGATTCCTACCAAGATGG	731
Db	609	TTGCTTCAGAAAGGTGCTAAATTTGGAGCCAACTTTTCGGTGGAAATGTTTCGACCAAGATGG	668
Qy	732	CTAGATTTTGGCGCAAGTTACAGCGCCCTGAGAGCTTTCGAGAGGACGGGAGGAATCTTGAC	791
Db	669	CTAGATTTTGTCTGAGTCTACAGCGCCCTGAGAGCCCTGAAGGACGGGAGAAACCAAGCC	728
Qy	792	GCTCCCAAGGTGCAAGAGTCTCTGAGTTACATCTGGGCTGGAGCTCTTCGCGGAAGACTCGC	851
Db	729	GCCCCCAAGCGCGAGAAATCTTTGAGTT- GCGTCAGCTCGACTCTCTC---AGACCCCTC	784
Qy	852	GTCTGTCACTACTCTTGTCAATTTGGCTTAGAGGAGGACTCCAGTGGGAGTCTGTCCTCAGA	911
Db	785	GGTCAACAGTCCCCCTTGATTTGGCTTAGAGGAGGGGATCCACTGGCAATACTTGGCCCCAGA	844
Qy	912	-CTGGAAGTTAAACTCAAGGCCCAAGAAAGAGCTTTAGACTCTCGACGGCCACTCTTCCT	970
Db	845	CCTAAATTTGAGCTTTAAGGCAAGGAGATGCTTTTGGACCTTGACACAGAGCTTTTCT	904
Qy	971	CCTGGAGCAGCAGCTGTGGGAGTGGAGTTGCTGCCCAGTAGCTTCAAAGCTGGTCTPAGT	1030
Db	905	CTTAGAGCAGCAGCTGTGGGAGTGGAGCTGTTGCCCCAGTAGGCTTCAATCCCGTCTGTA	964
Qy	1031	CTCCACCGAGAACTTGACTCTTTCATCCTCTGGGCCCTCTGAGCGTTCAGAGCTTAGGTAA	1090
Db	965	CTCTAAACCGGGAACTTGCTCTTCGCCCTCTGGGCCCTCTAAACATTTCAACGGCATAGACAA	1024
Qy	1091	TTTCAAGTAGTCTTCTATCTCTGAAACCTTCTCTACCTGACTACCTTCCCAAGTTAGG	1150
Db	1025	TTTCAGTGTGGTATTCCTATTTGCTTGAACCTTCTCTACCTGACTGCTTTCTTAGGCTAGA	1084
Qy	1151	GCTCGCTGTTCAGACGACGCTGGAGGTGGCCAGTTTGTGGGTTTCGGAACACTAAACCCC	1210
Db	1085	AGTCAGCTATCAGAACAAGTATGGAATAGCGAGGTAGTCGGCTTCAGACACTAACCCC	1144
Qy	1211	AGAGAGTGTCTATCTTTCTGAAGATGGTGTGTCAACCTCAGCGGTTGGGGCAGAGATGTC	1270
Db	1145	AGAGAGCAGCTGCCTGAGAGAGGACCAATTTGTATCCCGAGCGCTGAGTGCAGAACTCAT	1204
Qy	1271	GGCAACGCGCTGGAGAGGTGTCGCGCTCTCTCTACAGAGGCGTGGCGAATCCACCA	1330
Db	1205	TCCGGCCTCTGGCAGGGATGTCCACCTCTTTTACGGAAAGGCCCTACAGAAATTTACCA	1264
Qy	1331	CTTTCTGTATGAAAACGGCTAGAAATTTCTCT---CCAGGCTAAACAAAGGGCAAGAGTTACCCAC	1387
Db	1265	TCTTTCGATGAAAACGGCTGGAATTTCTTCAACAGGCTAGCAGGGGCAAGTTTACCCAC	1324
Qy	1388	CCCTGACCAAGATAATGGCTATCATAGCTGGAGGAGGAACATAA CTTTCTCGGATGGA	1447
Db	1325	CCCTGACCAAGATAATGCTACCAACAGCTGGAGGAGGAACACAGCCTTCTCGGATGGA	1384
Qy	1448	CCCAACAATTGCACAGATAAACCCAGCACAGGCGGTGTCCCTGCTGTGACACAGGCC---	1504
Db	1385	TCCAAAACACTGACAGAGATAACCCAAACACAGTTTGTCTGCTGTCTGGAGCATTTCTCGG	1444
Qy	1505	-----GGAGCCCACTGAGAAAAACCAAGAATGGTGTATTCAGAGTTTTCAC-----	1551
Db	1445	AAACACCCAGGAATCCACTTGAAGAAAAAATAGAAATTTAATACTACAGAGGTTCCACTTGC	1504
Qy	1552	-----AGAGCCCCCAGGGAAAGCAGTCTGTTTTGTGAAATTAGCCGTGGAAAAAGA	1600
Db	1505	TTTGGNAGNAGAGAGCCCTTCTGAGGCTGTCCATCTAGTGAGATACCTATGGAAAAAGGA	1564
Qy	1601	ATGTGAAGAGGACCACTAATAACTGAACTGACCTCTCAGATAGAGGAGAGAGCCTTCTGT	1660
Db	1565	GCCTGGAGAGGGCCGAAATAGTGTAGTTGATTATCATACCTAGAAAGTGACCTTCCCAT	1624

QY	1661	TTCTACACAGCAGCTTTGTGTAGCAACAACTGATAGATTATATTTTGGAGGCCGCCCCAG	1720
DB	1625	TTCTGCCAGACCCAGCTTGTAGTAAACAACTGATAGATTATATTTTGGAGGGTGATCCAG	1684
QY	1721	TGACTTGGAAAGCCAGCTCTGATTTCTGAAAGCTGAGGAATTTGGGGCGAGGAACCTCAGGACGA	1780
DB	1685	TGACTTGGAAACAAAGTTCTGATCCAGNAGGTGAGGATTTGGGATGAGGAAGCTGAGGATGA	1744
QY	1781	TGGCTTTGTATAGCGATGGCTCCCTGTCTGAAATCAGACGTGGAAACAGGACTCGGAAGGCCT	1840
DB	1745	TGGTTTGTATAGTGATAGCTCACTGTTCAGACTCAGACCTTGAACAAGACCTCTGAAGGCCT	1804
QY	1841	TCACCTTTGGAACTCTTTTCCACAGTGTAGATCCCTTACAAACCCCAAACTTTTACAGCCAC	1900
DB	1805	TCACCTTTGGAACCTCTTTCTGCAAGTGTAGATCCCTTAAATCCCCAGAACCTTTTACAGCAAC	1864
QY	1901	GATTACAGACGGCTGCCAGAATTGCCCCACAGACCCCATCAGATTTCAAGGACATCTCTGGTC	1960
DB	1865	AAATTCAGACTGCTGCCAGAAATTTGTTCTGGAAGAGCCTTCTGATTTCAGAGAAGGATTTGTC	1924
QY	1961	TGGCAGCTGTGTGTATGGGA---GCTGTTCAGGAGGAGCCCTTCCGGAGACCCCCGACCA	2017
DB	1925	TGGCAAGCTCGATCTAGAGAAATTCCTCCAGTCTGGAAGCCTTCTCGAGACCCCTCGACGA	1984
QY	2018	TAGTTCGGGGAGGAAGATCACTTGGGACCGAGTGCAGATGAAGCAGAGAACTCTTAAAT	2077
DB	1985	TAGTTCTGGGGAGGAAGATGACTGGGAATCTAGTGCAGATGAAGCAGAGAGTCTCAAACT	2044
QY	2078	GTGGAACTCTTCTGTCTCATTTCTGAGGACCCCTCAACCTTTTAAATTTTAAAGGCTCCTTT	2137
DB	2045	GTGGAACTCAATCTGTAAATTTCTGATGACCCCTCAACCCCTTAAATTTTAAAGGCTCCTTT	2104
QY	2138	TCAACCGCTCAGGGAGAAGTAATTTGGAAAGCCGCTCAGGACTCAAAAGGCCTCTTCTGAGGTAC	2197
DB	2105	TCAAACATCAGGGGAAATGAGAAAGGCTGCTCGTGACTCAGAGACCCCACTCTGAGTCCAT	2164
QY	2198	AGTGGCCTTCTCTGGCCATCATACCTTACTTTCTTGTAAAGCCGAGCTGTAGAGAGCCA	2257
DB	2165	TGTGGCCATTTCTGAGGTGCACACCTTACTTTCTTGTAAAGGTGCAGCTGTGTGGGAGCCA	2224
QY	2258	AGAAAGTAATTTGTCAGGCTGTGGCTGGGTGAGGCTCTTGTCTGGAGAAAGATACACCCA	2317
DB	2225	AGAAAGTGAATGTCCAGACTCGGTACAGCGTGACGTTCTTTCTTGGAGGAAGACACACACA	2284
QY	2318	TATCAAGAGAAAAAGGTAAACCTTCTCTGGAAGAGTTTACTGAGTATTATATAAGTGGTGA	2377
DB	2285	TGTCANAAAAAAGGTAAACCTTCTTCTTGAAGAGTTTACTGAGTATTATATAAGTGGTGA	2344
QY	2378	TGAGGATCGGAAGGACCATTTGGGAGAAATTTGCAAGGGATGGATGCAGGTTTCGAAACG	2437
DB	2345	TGAGGATCGGAAGGACCATTTGGGAGAAATTTTGAAGGGATGGATGCAGGTTTCCAGAAACG	2404
QY	2438	AAATTCAGAAACAGAGTTGCCATTGGCTACTCTGTGGCCTTTTGGACACAGAGAAAAAAT	2497
DB	2405	AAATTCAGAAACAGAGATGCTATTGGATATTGCTTGCATATTTTGAACACAGAGAAAGAAAT	2464
QY	2498	GTTTTAATAGACT-----GAGGATCGAGTCAAAGGACTTACTGTGTGTACAGCAATGTTAAGA	2553
DB	2465	GTTTTAATAGACTCCAGGGACATGCTTCAAAGGACTTAATGTTCTCAAGCAATGTTGAGT	2524
QY	2554	AGTGA-ACAGCTGCAACCCGTGCCACTCTGTCTCTTACTTGAGA-GTTTCCCTTAAAA	2611
DB	2525	TGGCAGGCTGTAGTCTCTAGCTAGCATACACTCTTACCTGAGAGGGTGCTCTTTTAAAAA	2584
QY	2612	ACAAACACTGGCAGCTGCTCTTGGACATGTTTTTAAAGAAACAACTTGTATCTAGAGATG	2671
DB	2585	ACAAATCTTGGCAGCTGCTCTTGGACATTTTTTTTTTTTATAGGAAATGTAACTTGGATCT	2644
QY	2672	CAGTTTGATATTTTTTGGGTAAATGTGTCTCATTTAGAA-----ACACCAACT	2717
DB	2645	AGTTTAATTTTTTTTTTTTGGCAATATCCCACTCAGAAACATTCAGGTTTGAAGCCAGCC	2704
QY	2718	CCGATAATGAAGAAAT-----CTCTTATCTGAATCCTCTCTTTTCTTATTTAGTTGGAT	2771

Db 2705 CTGATATGAAGGACTAGTGATTTCTAATCTCCCTTTTGTGATTAGTTGGAT 2764  
Qy 2772 GTGGGTTT 2779  
Db 2765 GTGCTTTT 2772

RESULT 13  
AL606489/c 167079 bp DNA linear PRI 08-NOV-2002  
LOCUS Human DNA sequence from clone RP11-739N20 on chromosome 1, complete  
DEFINITION  
ACCESSION AL606489  
VERSION AL606489.26 GI:24940211  
KEYWORDS HTG.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 167079)  
REFERENCE  
Pearce, A.  
Direct Submission  
Submitted (08-NOV-2002) Wellcome Trust Sanger Institute, Hinxton,  
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk  
On Nov 12, 2002 this sequence version replaced gi:22204251.  
----- Genome Center  
Center: Wellcome Trust Sanger Institute  
Center code: SC  
Web site: <http://www.sanger.ac.uk>  
Contact: humquery@sanger.ac.uk  
-----

During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.  
The following abbreviations are used to associate primary accession  
numbers given in the feature table with their source databases:  
Em, EMBL; Sw, SWISSPROT; Tr, TrEMBL; Wp, WORMPEP; Information  
on the WORMPEP database can be found at  
[http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) This sequence  
was generated from part of bacterial clone contigs of human  
chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping  
Group. Further information can be found at  
<http://www.sanger.ac.uk/HGP/Chri>  
RP11-739N20 is from the library RPCI-11.3 constructed by the group  
of Pieter de Jong. For further details see  
<http://www.chori.org/bacpac/home.htm>  
VECTOR: pBAC3.6

This sequence was finished as follows unless otherwise noted: all  
regions were either double-stranded or sequenced with an alternate  
chemistry or covered by high quality data (i.e., phred quality >=  
30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by at least  
one plasmid subclone or more than one M13 subclone; and the  
assembly was confirmed by restriction digest, except on the rare  
occasion of the clone being a YAC.

FEATURES  
Location/Qualifiers  
source  
1. 167079  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/clone="RP11-739N20"  
/clone\_lib="RPCI-11.3"

## ORIGIN

Query Match 20.6%; Score 1124.8; DB 9; Length 167079;  
Best Local Similarity 71.4%; Pred. No. 9e-273;  
Matches 1686; Conservative 0; Mismatches 607; Indels 67; Gaps 13;

Qy 39 CTGGGGCAATTTTGAGCTTCGCTTTCTTTGGCGCCCTCGCCCTGCCACCCAGCCACCCCTTTCC 98  
Db 63783 CTGGGGCAATTTTGGGCTTCGCTTC-----CACGGCACCGCGGCTTACCAGTCCCTTCC 63728  
Qy 99 GCCTTGGCGTTTCGGCGCTCCGTGCGGGCCACCGGAAACCGCGCGTGTCTCCGTGCGC 158  
Db 63727 GGTATCGCTTGTCTCAGGGGCTTTTCAACCTCTGTCTCAGTCGG-----AAAACCATCGCC 63673  
Qy 159 GCCGGCGAGGGAGGGTCTTCTATNGGTGGAGCGATCTCACAGGCGCTAGGACGTCTCC 218  
Db 63672 GAGGCGCTGGGGGAGCTCTTATCCATGGTGTGAAGCGTGGAGCCGACTAGGGAACCTCC 63613  
Qy 219 TTCCCTAGCCGGGATGACCTAACCGCGGTGCGCACCGCTTTGCGCGCGCTCTCTGGGCGGT 278  
Db 63612 TTCCC-CGCCAGATGGAAGTCCGATCAGTGC CGCTTATTTGCGGGGCTGTCTTCCCT 63554  
Qy 279 CCGGTGAGCAGTCTGTTTGGGAAAGCCCGCTCTCTGGGCTCTCTCTGCGGCGCGGAA 338  
Db 63553 GTGTTCTGCGCGCGCTGCGCATTCGCTGCGCTCTGTGGCTTTTCTGCTGGCTCGAAGA 63494  
Qy 339 TCGGACTGCGATACCCACTCGGTGGCTGGGCAAGGCGGAGACTGTGTAGACCTCGGATCC 398  
Db 63493 TCGGCTTGGAGCAGCGACGCCACCGCTGGGCAAGGCGGAGACTGTGTAGGCTTCTCCCGA 63434  
Qy 399 AGCTGCGCTGAGCCCGCTGAGCTCTCTCTCTCTCTCTGTCTGAGAAGCCGCAAGGA-AA 457  
Db 63433 ATCCCGTCGACCTCCAGCCCTGAGCGCCCGCGGCTTCTCTGAGAGACTGTCAAGAAAA 63374  
Qy 458 GGAGATGAGACAGGAAACGACAGGGCCCGGAAGCGGCTCTGGGCTCTGGCTGGGCTCTGT 517  
Db 63373 GGAGATGAGCGCGGACAGCGGATCGCGGAAACGGCTTGGGCTCTGGGCGGCTTTCCG 63314  
Qy 518 GTTCCGCTGCCC---TTCTCTGGGATCGACGCTCTCTTGGAGTTCGCCGCGCC 574  
Db 63313 GTTCTGGCCACCTTTTCTCTGGGATCGACGCTCTCTTCTCTGCCCAGCCGAGACTCG 63254  
Qy 575 TTCTCTCTGCAAAATCCCGGAA-----CTCGCTCTGCGCCGAGCGCTCG 619  
Db 63253 TCTTGGCCCGGAAATCCCGGAAACCCACCACTGCTTCTCTGCCCAGCCGAGACTCG 63194  
Qy 620 GACGAGTACTGAGACCAATTTGCTTTCTAGCTCTCTTGGCTCTCTCTCTAGCTATTCCA 679  
Db 63193 GGTCACTTCTGAGCAAACTGCTCTCCAGCTCTTGGCGCGCTCCCGGATTTGCTTCA 63134  
Qy 680 GAAGCTGCTCTTTGGAGCAGCTTCCGGGGGCTGATTCCTACCAAGATGGCTAGATT 739  
Db 63133 GAAGGTCTTAATTTGGAGCCAACCTTTTCGGTGGAAATGTTTCCGACCAAGATGGCTAGATT 63074  
Qy 740 TGCCGCAAGTTTACAGCCCTCTGAGAGCTTTCGAGAGGACGGGAGGAATCTGACGCTCCAC 799  
Db 63073 TGCTGGAGTCTACAGCCCTCTGAGAGCTTTCGAGAGGAGGAGAAACAGCGCCGCCAC 63014  
Qy 800 GGTGAGAGTCTCTGAGTTACTACTGGCTGGAGTCTTTCGCGAAGACTCGCGTCTGTCAG 859  
Db 63013 AGCGAGAAATCTTTGAGTT-CGCTGAGCTCGACTCTCTC---AGAGCCCTCGGTCACCA 62958  
Qy 860 TACTCTTGCATTTGGCTAGAGAGGACTCCAGTGGAGTGTCTGCTCTCAGA-CTGGAAG 918  
Db 62957 GTCCCTTGTATGGCTAGAGAGGGATTCACATGGCAATACTTCGCCCCAGACCTAAAT 62898  
Qy 919 TTAACCTCAAGGGCCAGGAAAGAGCTTTAGACTCTGAGCGGCCCACTTTCTCTCTGGAGC 978  
Db 62897 TGGAGCTTAAGGCCCAAGGAAGTCTTTGGACCTGCGAGCAGAGGCTTTCTCTTAGAGC 62838  
Qy 979 AGCAGCTGTGGGAGTGGAGTGTGCTGCCAGTAGCTCTTAAGCTGTGTCTAGTCTCCACC 1038  
Db 62837 AGCAGCTGTGGGAGTGGAGTGTGCTGCCAGTAGCTTCAATCCCGCTCTGTACTCTAAC 62778  
Qy 1039 GAGAACTTGACTCTTCACTCTGCGGCTCTGAGCGTTCAGAGCTTAGGTAAATTTCAAG 1098  
Db 62777 GGGAACTTGGCTCTTCCCTCTGGGCTCTTAACATTTCAACGATAGACAAATTTAGTG 62718

```
Qy 1099 TAGTTTCCTATCTCTGAAACCCCTTCCTACCTGGAGTACCTTCCCAAGTTTAGGGCTGGCT 1158
Db 62717 TGGTATCCTATTTGCTGAAACCCCTTCCTACCTGGAGTACCTTCCCTAGGCTAGAGTCAAGCT 62658
Qy 1159 GTACAGCAGCGCTGGAGGTGGCCAGTTTGTGGGTTCGGAACACTAACCCCGAGAGCT 1218
Db 62657 ATCAGAACAGTGATGGAATAGCAGGTAGTCGGCTTCCAGACACTAACCCCGAGAGCA 62598
Qy 1219 GCTATCTTTCTGAAGATGGTTGTCAACCTCAGCCGTTTGGGGGCGAGAGATGTGGCAACCG 1278
Db 62597 GCTGCTTGAGAGAGGACCATTTGTCAATCCAGCGCTGAGTGCAAGAACTCATTCGGGCT 62538
Qy 1279 CTTGGAGAGGTGTCCGCTCTCTACAGAGGCTTCCGGAATCCACACCTTCGTA 1338
Db 62537 CGTGGCAGGATGTCCACCTCTTTCTACGGAAGGCTTACAGAAATTCACCATCTTCGCA 62478
Qy 1339 TGAACCGCTAGAAATTCCT---CCAGGCTAACAAAGGCAAGAGTTACCCACCCCTGCAC 1395
Db 62477 TGAACCGCTGGAATTCCTCAACAGGCTAACAAAGGCGCAAGATTTACCCACCCCTGCAC 62418
Qy 1396 AAGATAATGGCTATCATAGCTGGAGGAGGAACAATAACCTTCTCCGGATGGACCCACAAC 1455
Db 62417 AGGATATGGCTACACAGCTCGAGGAGGAACAACAGCCTTCTCCGGATGGATCCAAAC 62358
Qy 1456 ATTGCACAGATAACCCAGCACAGCGGTGTCCTCTGTCGACAGCAGGC----- 1504
Db 62357 ACTGACAGATAACCCACAACAGTTGTGTCGCTGCTGGAGACATTCCTGGAAACACCC 62298
Qy 1505 -GGAGCCACTGAGAAAACCAATTTGGTGAATTCAGAGTTTCAC----- 1551
Db 62297 AGGAATCCACTGAAAGAAAATAAGATTAATTAATACAGAGTTCCACTTGTCTTGGAA 62238
Qy 1552 ---AGAGCCCCCAGGGAAGCAGTCTGTTTGTGAATTAACCGTGGAAAAAGAAATGTGAAG 1608
Db 62237 AAGAGAGCCCTTCGTGAGGCTGTCCATCTAGTAGATACCTATGGAAGAGGAGCCTGGAG 62178
Qy 1609 AGGACCACTAATATGCAATGCACTCTCAGATAGAGGAGAGAGCCTTCTGTTTCTACCA 1668
Db 62177 AGGCGCAATTAAGTGTAGTTGATTACTCATACCTAGAAAGGTGACCTTCCCATTTCTGCCA 62118
Qy 1669 GACCAAGTTGTAGCAACAACTGATAGATTAATTTTGGAGGCGCCCCCAGTACACTGG 1728
Db 62117 GACCAGCTTGTAGTAACAACTGATAGATTAATTTTGGAGGTGCAATCCAGTACCTGG 62058
Qy 1729 AAGCCAGCTCTGATTTCTGAAAGTGAAGATTGGGCGGAGGAACCTCGAGACGATGGCTTTG 1788
Db 62057 AAACAAGTTCTGATCCAGAGGTGAGGATTTGGGATGAGAGCTGAGATGATGGTTTG 61998
Qy 1789 ATAGCGATGGCTCCCTGTGTAATCAGACGTGGGAACAGGACTCGGAAGGCTTCCACCTTT 1848
Db 61997 ATAGTGATAGCTCACTGTGCAGCTCAGACCTTGAACAAGACCTGGAAGGCTTCCACCTTT 61938
Qy 1849 GGAATCTTTCCAGTGTAGATCTTTACAAACCCCAAACTTTACAGCCAGATTCAGA 1908
Db 61937 GGAATCTTTCTGCAGTGTAGATCTTTAATATCCCAAGAACTTTACAGCAACAATTCAGA 61878
Qy 1909 CGGCTGCCAGAAATTTGCCCCAGAGACCATCAGATTCAGGACATCTCTGGTCTGGCAGCT 1968
Db 61877 CTGCTGCCAGAAATTTGTTCTCCTGAAGAGCTTCTGATTCAGAGAGGATTTGTCTGGCAGT 61818
Qy 1969 GTGTTGTAGGGA---GCTGTGAGGAGGACCCCTTCCGGAGACCCCGACCATAGTTCCG 2025
Db 61817 CTGATCTAGAGAAATTCCTCCAGCTGGAAGCCTTCTCTGAGACCCCTGAGCATAGTTCTG 61758
Qy 2026 GGGAGGAGATGACTGGGAACCGAGTGAGATGAAGCGAGAAATCTTAATTTGGAACT 2085
Db 61757 GGGAGGAGATGACTGGGAATCTAGTGAGATGAAGCGAGAGAGTCTCAAACTGTGGAACT 61698
Qy 2086 CTTTCTGTCTTCTGAGAGCCCTTACACCTTTTAAATTTTAAAGGCTCTTTTCAACCGT 2145
Db 61697 CATTTCTGTAATCTCTGATGACCCCTTACACCCCTTTAAATTTTAAAGGCTCTTTTCAACAT 61638
Qy 2146 CAGGGAAGAAATTTGGAAGGCGCTCAGGACTCAAAAGGCGCTTCTTCTGAGGTCAAGTGGCCT 2205
```

```
Db 61637 CAGGGGAAAATGAGAAGGCTGCTGACTCAAAGACCCCATCTGAGTCCATTGTGGCCA 61578
Qy 2206 TCTCTGGCCATCATACCTTACTTCTTGTAAAGCCCGAGCTGTTAGAGAGCCCAAGAAAGATA 2265
Db 61577 TTTCTGAGTGTCAACCTTACTTCTTGTAAAGTGCAGCTGTTGGGGAGCCCAAGAAAGTG 61518
Qy 2266 ATTGTCCAGGCTGTGGGCTGGGTGAGGCTTCTGAGAGAAAGATACACCCATATCAAGA 2325
Db 61517 AATGTCAGACTCGGTACAGCGTGAAGCTTCTTCTGAGGAGACACACACATGTCAAAA 61458
Qy 2326 GAAAAAGGTAACTTCCCTG 2345
Db 61457 GAAAAAGGTTTGTCTTCTG 61438

RESULT 14
AL627202/c
LOCUS Homo sapiens chromosome 1 clone RP11-174P23, 2 unordered pieces.
DEFINITION AL627202
ACCESSION AL627202
VERSION 6 GI:24394911
KEYWORDS HTG; HTGS PHASE1; HTGS_CANCELLED.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 328863)
AUTHORS Clark, S.
TITLE Direct Submission
JOURNAL Submitted (21-OCT-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Oct 25, 2002 this sequence version replaced gi:20268846.
COMMENT ----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA174P23
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 170058 bases at least Q40
Consensus quality: 171825 bases at least Q30
Consensus quality: 172641 bases at least Q20
Insert size: 328763; sum-of-contigs
Quality coverage: 172230; 12.6% error; agarose-fp
Quality coverage: 2.67x in Q20 bases; sum-of-contigs Quality
coverage: 5.15x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 6943: contig of 6943 bp in length
* 6944 7043: gap of 100 bp
* 7044 328863: contig of 321820 bp in length.

FEATURES
source
1..328863
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-174P23"
/clone_lib="RPCI-11.1"
misc_feature 1..6943
```

```
/note="assembly fragment:03428"
```

7044. 328863

```

/note="assembly fragment:03558"

```

misc feature

## ORIGIN

Query Match	20.6%	Score 1124.8	DB 2	Length 388863
Best Local Similarity	71.4%	Pred. No. 1.1e-272		
Matches 1686	Conservative 0	Mismatches 607	Indels 67	Gaps 13
Qy	39	CTGCGGCCATTTCGAGCTTCGCTCTTTGTCGCGCCCTCGCTGCGACCCAGCCACACCTTTC	98	
Db	225570	CTGCGGCCATTTCGCGCTTCGCTTC-----CACCGCACAGCGCGCTTACCAGTCTCTCC	225515	
Qy	99	GCCTTGGCGTTTCGCGCTCCGTCGGGCGCACCGGAAACGCGCCGCTGCTTCGCGTCGCC	158	
Db	225514	GGTATCGCTTGTCTCAGGGCTTTTCAACCCCTCTGTCACTCG-----AAACCATCGCC	225460	
Qy	159	GCCCGCGAGGAGGGTCTTCTATATGTTGAGCGATCTCACAGGCTTAGAGCTCTCC	218	
Db	225459	GAGGCCGTGGGGGACTCTTATCATGTTGTGAAGCGTCGAGCGGACTTAGGGAACCTCC	225400	
Qy	219	TTCCCTAGCCGGGATGAGACCTAAACCGCGGTCGCCACCGCTTTCGCGGGCGCTTCGGCCGT	278	
Db	225399	TTCCC-GCCAGGATGNAATGTCGATCAGTCGCGCCCTATTTCGCGGGCTGTTCTTCCT	225341	
Qy	279	CCGCTGAGCACTCGTTTCGGAAACCGCGCTCTCTCTGGGCTCTCTGTCGCGCGCGGAA	338	
Db	225340	GTGTCTGCGCGCGCTGCCCATTCGCTGCGCTCTGTGGCTTTTCTGCTGGCTCGAAGA	225281	
Qy	339	TGCGACTCGAGTACCCACTCCGTGGCTGGGCAAGCGGAGACTGTGTAGACTCTCGGATCC	398	
Db	225280	TCGGCCTTGGAGACGACGACGCCACCGCTGGGCAAGCGCGAGACTCTGTAGGCTTCTCCGA	225221	
Qy	399	AGCTCGCGCTGACGCCCGCTGAGCTCTGTCTCTCTCTGTCTGAGAAAGCGCGCAAGAA	457	
Db	225220	ATCCGTCGACCTCCAGCGCTGAGCGCGCGGCCCTACCTGAGAGACTGTCAAGAAAA	225161	
Qy	458	GGAGATGGAGACAGGAACGACACAGGCGCCGGAAGCGGCTTGGCCCTCGGCTCGGCTCTG	517	
Db	225160	GGAGATGGAGCCGGGACAGCGCGATCGCGAAACGCGTTTGGCCCTCGGCGGGCTTCG	225101	
Qy	518	GTTCGCGCTGCC--TTCTCTTCGGCGATCGACGCTGCTCTTCGGAGTTCCGCGCGCC	574	
Db	225100	GTTCGCGCACCTTTTTCCTCGCGGATCGAAGAGGCTCTTCTAAGTTTCCGACGCC	225041	
Qy	575	TTCTCTCGACAAATCCCGGAA-----CTCCGCTCTGCCGCGCGCTCG	619	
Db	225040	TCTTGGCCCGGAAACTCCGGGAACCCACACTGCTTTCTCTGCCCCAGCCCGAGACTCG	224981	
Qy	620	GACCAGTACTGGACCAATTTGCTTTCTCAGCTCTTGGCCCTGCTCCCTAGCCTATTCCA	679	
Db	224980	GGTCAGTTACTGGACGAATCTCTCCAGCTCTTTCGCGCGCTCCCGGATTCGTTCA	224921	
Qy	680	GAAGCTCTGCTTTGGAGCAGCTTTTCGGGGGCTGATTCCTACCAGATGGCTAGATT	739	
Db	224920	GAAGTGCTAAATTTGGAGCCAACTTTTTCGGTGGAAATGTTTCCGACAGATGGCTAGATT	224861	
Qy	740	TGCGCAAGTTACACGCCCTTGAGACTTCGAGAGGACGGGAGAAATCTGACGCTCCAC	799	
Db	224860	TGCTGGAGTCTACACGCCCTTGAGACCTTGAAGGCGGAGAAACGAGCGCCCCC	224801	
Qy	800	GGTGCAGAGCTCTGAGTTACATGCGGCTGGAGCTTTCGGAAGACTCGGCTCGTCAG	859	
Db	224800	AGGCGAATAATTTTGAAGTT-CGCTGAGCTTCGACTCTCTC-----AGACCCCTCGTCA	224745	
Qy	860	TACTCTTGCATTGGCTAGAGAGGGAATCCAGTGGCAGTGTCTGCTCTCAGA-CTGGAAG	918	
Db	224744	GTCCCTTGATTGGCTAGAGAGGGATCCATGGCAATACTCGCCCCAGACCTAAAT	224685	
Qy	919	TTAAACTCAAGCCCGAGAAAGACTTTAGACTCTGACAGCGCCCACTTTCTCTCGAGC	978	
Db	224684	TGAGCTTAAAGCCAGGGAAGTGCTTTGGACCTTGACGACAGGCTTTCTCTTAGAGC	224625	

```

Db 223544 GGGAGGAAGTCTAGTGGGAATCTAGTCAGATGAAGCAGAGAGTCTCAAACTGTGGAACT 223485
Qy 2086 CTTTCTGCTGCTTCTGAGGACCCCTACCAACCTTTTAAATTTTAAAGGCTCCTTTTCAACCGT 2145
Db 223484 CATTCCTGAATCTGATGACCCCTACCAACCTTTTAAATTTTAAAGGCTCCTTTTCAACAT 223425
Qy 2146 CAGGGAAGAAATTGGAAGGCGCTCAGGACTCAAGGCGCTCTTCTGAGGTCAACAGTGGCCT 2205
Db 223424 CAGGGAAGAAATGAGAAAGGCTGTCGTGACTCAAGAGACCCCATCTGAGTCCATTGTGGCCA 223365
Qy 2206 TCTCTGCCCATCATACCTTACTTCTTGTAAAGGCCAGCTGTTAGAGAGGCCAAGAATA 2265
Db 223364 TTTCTGAGTGTCAACACTTACTTCTTGTAAAGTGCAGCTGTTGTGGGAGGCCAAGAAAGTG 223305
Qy 2266 ATTGTCCAGGCTGTGGGCTGGGTGAGGCTCTTGTCTGAGAAAGATACACCATCATCAAGA 2325
Db 223304 ATGTCCAGACTCGGTACAGGTGACGTTCTTCTGAGGAGACACACATGTCAAAA 223245
Qy 2326 GAAAAAGGTAACTTCTCTG 2345
Db 223244 GAAAAAGGTTGTTCTTG 223225

RESULT 15
AC048369 166233 bp DNA linear HTG 25-JUN-2000
DEFINITION Homo sapiens chromosome 1 clone RP11-284G5 map 1, WORKING DRAFT
SEQUENCE 29 unordered pieces.
AC048369
VERSION AC048369.2 GI:8705131
HTG; HTGS PHASE1; HTGS_DRAFT.
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abram,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Basien,V., Beda,F.,
Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G.,
Campiano,A., Castle,A., Choquel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Mienga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Teefaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo.A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (14-APR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jun 25, 2000 this sequence version replaced gi:7549699.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
-----Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu

```

```

Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L8045
Center clone name: 284_G_5
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 151910 bases at least Q40
Consensus quality: 158917 bases at least Q30
Consensus quality: 161704 bases at least Q20
Insert size: 178000; agarose-fp
Insert size: 163433; sum-of-ctnigs
Quality coverage: 3.7 in Q20 bases; agarose-fp
Quality coverage: 4.0 in Q20 bases; sum-of-ctnigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 29 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1069: contig of 1069 bp in length
* 1070 1169: gap of 100 bp
* 1170 2738: contig of 1569 bp in length
* 2739 2838: gap of 100 bp
* 2839 4143: contig of 1305 bp in length
* 4144 4243: gap of 100 bp
* 4244 5489: contig of 1246 bp in length
* 5490 5590: gap of 100 bp
* 5591 6873: contig of 1284 bp in length
* 6874 9153: contig of 2180 bp in length
* 9154 9253: gap of 100 bp
* 9254 10555: contig of 1302 bp in length
* 10556 10655: gap of 100 bp
* 10656 13995: contig of 3340 bp in length
* 13996 14095: gap of 100 bp
* 14096 16689: contig of 2594 bp in length
* 16690 16789: gap of 100 bp
* 16790 19295: contig of 2506 bp in length
* 19296 19395: gap of 100 bp
* 19396 22797: contig of 3402 bp in length
* 22798 22897: gap of 100 bp
* 22898 26453: contig of 3556 bp in length
* 26454 26553: gap of 100 bp
* 26554 31369: contig of 4816 bp in length
* 31370 31469: gap of 100 bp
* 31470 35528: contig of 4059 bp in length
* 35529 35628: gap of 100 bp
* 35629 40531: contig of 4903 bp in length
* 40532 40631: gap of 100 bp
* 40632 45154: contig of 4523 bp in length
* 45155 45254: gap of 100 bp
* 45255 50400: contig of 5146 bp in length
* 50401 50500: gap of 100 bp
* 50501 56357: contig of 5857 bp in length
* 56358 56457: gap of 100 bp
* 56458 62667: contig of 6210 bp in length
* 62668 62767: gap of 100 bp
* 62768 69282: contig of 6515 bp in length
* 69283 76270: gap of 100 bp
* 76271 76370: contig of 6888 bp in length
* 76371 84430: contig of 8060 bp in length
* 84431 84531: gap of 100 bp
* 84532 92060: contig of 7530 bp in length
* 92061 92160: gap of 100 bp
* 92161 101111: contig of 8951 bp in length
* 101112 101212: gap of 100 bp
* 101213 109009: contig of 7798 bp in length

```







Db	3735	AGCCTGGAGAGGCGCGAATAAGTGTAGTTGATTACTCATACCTAGAAGGTGACCTTCCCA	3794
Qy	1660	TTTCTACAGACACAGTTTGTAGCAACAACTGATAGATTATATTTTGGGAGGCGCCCCCA	1719
Db	3795	TTTCTGCCAGACCAAGCTTGTAGTAACAACCTGATAGATTATATTTTGGGAGGTGCATCCA	3854
Qy	1720	GTGACTTGGAAAGCCAGCTCTGATTCTGAAAGTGAAGGATTGGGGCGAGGAACCTTGAGGACG	1779
Db	3855	GTGACCTGGAAACAAGTTCTGATCCAGAAGGTGAGGATTGGGATGAGGAAGCTGAGGATG	3914
Qy	1780	ATGCTTTTGATAGCGATGGCTCCCTCTGCTGAATCAGACGTGGAACAGGACTCGGAAGGCC	1839
Db	3915	ATGGTTTGTAGTAGTAGCTCACTGTGAGACTCAGACCTTGAACAAGACCTGAAGGGC	3974
Qy	1840	TTACACCTTTGGAACCTCTTCCACAGTGTAGATCCTTACAAACCCCAAACTTTACAGCCA	1899
Db	3975	TTACACCTTTGGAACCTCTTCTGCAGTGTAGATCCTTATATCCCGAGAACCTTTACAGCAA	4034
Qy	1900	CGATTACAGACGGCTGCCAGAAATGCCCCACAGAGACCCATCAGATTCAGGGACATCCTGGT	1959
Db	4035	CAATTACAGACTGCTGCCAGAAATTTCTCTGAAGAGCCTTCTGATTCAGAGAAGGATTGT	4094
Qy	1960	CTGCCAGCTGTGTTGTAGGGA	1980
Db	4095	CTGGCAAGTCTGATCTAGAGA	4115

Search completed: September 16, 2005, 01:58:03  
 Job time : 22799.5 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 15, 2005, 08:08:59 ; Search time 2648.83 Seconds  
(without alignment)  
12220.183 Million cell updates/sec

Title: US-10-650-482-3

Perfect score: 5468

Sequence: 1 cggctcctcgctcgccctg.....attaaaaaaaaaaaaaa 5468

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 100 summaries

Database : N\_Geneseq\_16Dec04:\*  
1: Geneseq1980s:\*  
2: Geneseq1990s:\*  
3: Geneseq2000s:\*  
4: Geneseq2001as:\*  
5: Geneseq2001bs:\*  
6: Geneseq2002as:\*  
7: Geneseq2002bs:\*  
8: Geneseq2003as:\*  
9: Geneseq2003bs:\*  
10: Geneseq2003cs:\*  
11: Geneseq2003ds:\*  
12: Geneseq2004as:\*  
13: Geneseq2004bs:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5468	100.0	5468	12	ADQ88459 Mouse GAD
2	1324	24.2	2942	12	ADQ88457 Human GAD
3	1320.8	24.2	2942	4	AH18699 Human cDN
4	1189.2	21.7	2324	3	AAH98158 Human col
5	542.6	9.9	883	4	AAH99389 Human pro
6	239.2	4.4	772	4	AAH06623 Human cDN
7	181.6	3.3	540	6	ABK63389 Rat seque
8	181.6	3.3	540	10	ADBS57750
9	170.8	3.1	399	8	ABX52167 Bovine ES
10	113.6	2.1	689	4	AAI96694 Human neu
11	109	2.0	421	9	ACH49151 Human leu
12	109	2.0	543	4	AAH11666 Human cDN
13	99.6	1.8	1028	3	AAF21761 Human bre
14	98	1.8	482	4	AAH13104 Human cDN
15	98	1.8	2148	4	AAH16576 Human cDN
16	98	1.8	2148	6	ABV77954 Hypoxia-r
17	98	1.8	2148	10	ADG32888 Human DNA
18	93.8	1.7	218	3	AC229803 Human sec
19	85.8	1.6	10610	12	ADQ97608 Mouse can
20	84	1.5	1678	2	AAT34611 Probe for

c	21	82.6	1.5	90141	12	ADQ97867	Adq97867 Mouse can
c	22	79.4	1.5	8351	6	ABK63492	Abk63492 Rat seque
c	23	79.4	1.5	8351	10	ABT41759	Abt41759 Toxicity
c	24	79.4	1.5	8351	12	ADP72626	Adp72626 Renal tox
c	25	79.4	1.5	119414	11	ACN45052	Acn45052 Mouse gen
c	26	79.2	1.4	1448	10	ADBS8896	Adbs8896 Toxicity-
c	27	79.2	1.4	1448	10	ADBS3612	Adbs3612 Primary r
c	28	79.2	1.4	1448	10	ADBS3612	Adbs3612 Primary r
c	29	79.2	1.4	1454	3	AAC97642	Aac97642 Rat Bridg
c	30	79.2	1.4	4529	10	ADBS8829	Adbs8829 Toxicity-
c	31	79.2	1.4	4529	10	ADBS3548	Adbs3548 Primary r
c	32	79.2	1.4	4529	10	ADBS3548	Adbs3548 Primary r
c	33	79.2	1.4	25891	11	ACN44952	Acn44952 Mouse gen
c	34	79.2	1.4	53963	9	ADA02537	Ada02537 Mouse Tco
c	35	79.2	1.4	53963	10	ADBS72275	Adbs72275 Mouse Tco
c	36	79.2	1.4	53963	10	ADBS95785	Adbs95785 Mouse Tco
c	37	78.8	1.4	3046	2	AAV29074	Aav29074 Rattus no
c	38	78.6	1.4	1496	6	ABL34990	Ab134990 Rat cDNA
c	39	78.6	1.4	13894	10	ADBS59134	Adbs59134 Toxicity-
c	40	78.2	1.4	71132	11	ACN45092	Acn45092 Mouse gen
c	41	77.8	1.4	1432	10	ADBS58133	Adbs58133 Toxicity-
c	42	77.8	1.4	1432	10	ADBS52618	Adbs52618 Primary r
c	43	77.6	1.4	5531	10	ADBS3403	Adbs3403 Primary r
c	44	77.6	1.4	5531	10	ADBS42272	Adbs42272 Toxicity
c	45	77.6	1.4	215980	6	AAL38337	Aal38337 Complemen
c	46	77.4	1.4	433	12	ADI04042	Adi04042 Human neu
c	47	77.4	1.4	233060	11	ACN43912	Acn43912 Mouse gen
c	48	77	1.4	36119	13	ABD33605	Abd33605 Murine ca
c	49	76.6	1.4	6251	6	ABK62565	Abk62565 Rat seque
c	50	76.6	1.4	6251	10	ABT40825	Abt40825 Toxicity
c	51	76.6	1.4	6251	12	ADP71847	Adp71847 Renal tox
c	52	76	1.4	478	10	ADBS1015	Adbs1015 Primary r
c	53	76	1.4	479	10	ADBS1404	Adbs1404 Primary r
c	54	76	1.4	575	10	ADBS7185	Adbs7185 Toxicity-
c	55	76	1.4	120239	13	ABD33244	Abd33244 Murine ca
c	56	75.4	1.4	508	6	ABK62412	Abk62412 Rat seque
c	57	75.2	1.4	2070	9	ACF36089	Acf36089 Rat FXR p
c	58	75.2	1.4	2070	9	ACF04415	Acf04415 Rat farne
c	59	75.2	1.4	2070	10	ADBS58407	Adbs58407 Toxicity-
c	60	75.2	1.4	2070	10	ADBS52987	Adbs52987 Primary r
c	61	75.2	1.4	2070	10	ABT42054	Abt42054 Toxicity
c	62	75.2	1.4	2070	12	ADP72840	Adp72840 Renal tox
c	63	75.2	1.4	3122	5	AA91297	Aa91297 Rat Urote
c	64	75.2	1.4	3122	13	ADM99718	Adm99718 Norway ra
c	65	75	1.4	2472	10	ADBS58312	Adbs58312 Toxicity-
c	66	75	1.4	2472	10	ADBS52866	Adbs52866 Primary r
c	67	75	1.4	2472	12	ADJ10118	Adj10118 Rat vitam
c	68	75	1.4	2485	10	ADBS3252	Adbs3252 Primary r
c	69	75	1.4	2485	10	ADBS42199	Adbs42199 Toxicity
c	70	74.8	1.4	2010	6	AAH94721	Aah94721 Rat seque
c	71	74.6	1.4	227	5	AAH82230	Aah82230 Rat diffe
c	72	74.4	1.4	10929	13	ADT66569	Adt66569 Rat Lutzp
c	73	74.2	1.4	13011	2	AAT96631	Aat96631 cDNA enco
c	74	74.2	1.4	13011	10	ABT42448	Abt42448 Toxicity
c	75	74.2	1.4	13011	12	ADP72914	Adp72914 Renal tox
c	76	74	1.4	411	12	ADP72375	Adp72375 Renal tox
c	77	74	1.4	507	12	ADP71796	Adp71796 Renal tox
c	78	74	1.4	3134	6	ABL34973	Ab134973 Murine cD
c	79	73.6	1.3	1125	12	ADQ35602	Ado35602 Novel mou
c	80	73.6	1.3	7577	13	ABD32996	Abd32996 Mouse can
c	81	73.6	1.3	7997	2	AAQ89553	Aaq89553 Rat chole
c	82	73.6	1.3	7997	2	AAT79737	Aat79737 Rat chole
c	83	73.6	1.3	109147	12	ADQ97718	Adq97718 Mouse can
c	84	73.4	1.3	427	2	AAT34599	Aat34599 Probe for
c	85	73	1.3	38239	12	ADQ97626	Adq97626 Mouse can
c	86	73	1.3	109565	13	ABD33086	Abd33086 Murine ca
c	87	72.8	1.3	1089	10	ADCS3443	Adcc3443 Rat pl8Ab
c	88	72.8	1.3	1665	3	AAA90869	Aaa90869 Rat stres
c	89	72.8	1.3	1686	10	ADBS53061	Adbs53061 Primary r
c	90	72.8	1.3	2006	10	ADBS58134	Adbs58134 Toxicity-
c	91	72.8	1.3	2006	10	ADBS2619	Adbs2619 Primary r
c	92	72.8	1.3	2006	10	ADBS2619	Adbs2619 Primary r
c	93	72.8	1.3	2279	10	ADBS6187	Adbs6187 Toxicity-

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360	361	362	363	364	365	366	367	368	369	370	371	372	373	374	375	376	377	378	379	380	381	382	383	384	385	386	387	388	389	390	391	392	393	394	395	396	397	398	399	400	401	402	403	404	405	406	407	408	409	410	411	412	413	414	415	416	417	418	419	420	421	422	423	424	425	426	427	428	429	430	431	432	433	434	435	436	437	438	439	440	441	442	443	444	445	446	447	448	449	450	451	452	453	454	455	456	457	458	459	460	461	462	463	464	465	466	467	468	469	470	471	472	473	474	475	476	477	478	479	480	481	482	483	484	485	486	487	488	489	490	491	492	493	494	495	496	497	498	499	500	501	502	503	504	505	506	507	508	509	510	511	512	513	514	515	516	517	518	519	520	521	522	523	52
--	---	---	---	---	---	---	---	---	---	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	----

## ALIGNMENTS

RESIST 1

AD088459

ID ADO88459 standard: cDNA: 5468 bp.

AC AD088459.

07-OCT-2004 (first entry)

XX  
DE  
Mouse GADD34-like (GADD34L) cDNA.

AA  
KW Oxidative stress; neuronal ischaemia; heart ischaemia; renal damage;  
KW autoimmune disease; neurodegenerative disorder; therapy; GADD34-like;  
KW GADD34L; eIF2alpha-specific regulatory subunit; phosphatase; mouse; gene;

XX

XX  
XX

— 2 —

FH	Key	Location/Qualifiers
----	-----	---------------------

FT	CDS	FT	CDS
462.	.2558	462.	.2558

FT /\*tag= a

FT /product= "Mouse GADD34-like (GADD34L) protein"

```
FT /transl_except= (pos:1332. .1334, aa:Arg)
```

```
FT = FT + 1
    /transl_except= (pos:1341. .1343, aa:Arg)
```

```
FT      /transl_except= (pos:1344. .1346, aa:Trp)
```

```
FT      /transl_except= (pos:1350. .1352, aa:Val)
EM      /transl_except= (pos:1350. .1354, aa:Val)
```

```

FT      /transl_except= (pos:1362: .1364, aa:Pro)
DM      /transl_except= (pos:1368: .1370, aa:Gln)

```

```

F-I /transl_except= (pos:1368. .1370, aa:GIN)
PT  /transl_except= (pos:1377 1378 aa:ACG)

```

```
FT      /tgrant_except= (row:1377. .1379, aa:avb)
PT      /tgrant_except= (row:1380 .1381 aa:avb)
PT      /tgrant_except= (row:1382 .1383 aa:avb)
```

```

FI
XX
/СТАВЫ_ЕХСЕРС= (Р08:1309. .1391, аа:1еи)

```

PN IIS2004142345-A1

22-111.-2004

28-AUG-2003. 2003UIS-00650482

06-SEP-2002. 2002US-0408678D

XX  
DA (BOND) / BOND D

PA (KOND//) KON D.  
PA (.TOIS/) TOISSE C

XX  
PT  
Bon D  
Toussan C.

XX	WT	CCC/C
22	004	CCC/C

DR WPI; 2004-552556/  
R-PCNB. 1D089460

xx PT Screening test substances for preventing or treating disease involving  
PT oxidative stress, by testing test substances for its ability to inhibit  
PT activity of GADD34L and identifying test substance that inhibits activity  
PT of GADD34L.

PS Disclosure: SEO ID NO 3: 30pp: English.

The present invention relates to a method of screening several test substances for preventing or treating diseases involving oxidative stress such as neuronal ischaemia, heart ischaemia, renal damage induced by ischaemia or toxins, autoimmune diseases and neurodegenerative disorders. The method involves testing the test substances for its ability to inhibit the activity of GADD34-like (GADD34L), also referred to as eIF2alpha-specific regulatory subunit of phosphatase, and identifying the test substance which inhibits the activity of GADD34L. The present

CC	sequence	is mouse	GADD34L	cDNA.
XX				
SQ	Sequence	5468 BP; 1321 A; 1272 C; 1366 G; 1509 T; 0 U; 0 Other;		
	Query Match	100.0%; Score 5468; DB 12; Length 5468;		
	Best Local Similarity	100.0%; Pred. No. 0;		
	Matches 5468; Conservative	0; Mismatches 0; Indels 0; Gaps 0;		
Qy	1	CGGTCTCTCCGTCTCGCCCTCAGCTTCGCGGTGTGCGGCTGCGGCCATTGTGAGCTTCGC	60	
Db	1	CGGTCTCTCCGTCTCGCCCTCAGCTTCGCGGTGTGCGGCTGCGGCCATTGTGAGCTTCGC	60	
Qy	61	TTCTTTGGGCCCTCGCTCGCCACCGACGACCCCTTTCGGCTTGGCGCTTTCGGCGCTCCG	120	
Db	61	TTCTTTGGGCCCTCGCTCGCCACCGACGACCCCTTTCGGCTTGGCGCTTTCGGCGCTCCG	120	
Qy	121	TGCGGGCCACCGGAAAGCGCGCGTCTCGTCCGTCGCGCGCGCGCGAGGAGAGGCTCTTCT	180	
Db	121	TGCGGGCCACCGGAAAGCGCGCGTCTCGTCCGTCGCGCGCGCGAGGAGAGGCTCTTCT	180	
Qy	181	CTATGTGTGAGCGATCTTCACACGGCCCTAGGACGCTCTCTCCCTAGCCGGGATGGACCTA	240	
Db	181	CTATGTGTGAGCGATCTTCACACGGCCCTAGGACGCTCTCTCCCTAGCCGGGATGGACCTA	240	
Qy	241	ACCGCGGTTCGCCACCGCTTGC	300	
Db	241	ACCGCGGTTCGCCACCGCTTGC	300	
Qy	301	AGCGCGGCTCTCTGGGCTCTCTGCGCGCGCGGGAATCGGACTCGAGTACCCACTCCG	360	
Db	301	AGCGCGGCTCTCTGGGCTCTCTGCGCGCGCGGGAATCGGACTCGAGTACCCACTCCG	360	
Qy	361	TGGCTGGCAAGCGGAGACTGTGTAGACTCTCGGATCCAGCTCGCTGACGCGCTGAG	420	
Db	361	TGGCTGGCAAGCGGAGACTGTGTAGACTCTCGGATCCAGCTCGCTGACGCGCTGAG	420	
Qy	421	CTCTGTCTCTCTCTGTCTGAGAGCGCGCAAGGAAGGAGATGGAGACAGGAACGCACA	480	
Db	421	CTCTGTCTCTCTCTGTCTGAGAGCGCGCAAGGAAGGAGATGGAGACAGGAACGCACA	480	
Qy	481	GGGCGCGGAAGCGGCTGGCCCTCGGCTGGGCTCTGGGTTCGGCTGCGGCTTCTTTCGGC	540	
Db	481	GGGCGCGGAAGCGGCTGGCCCTCGGCTGGGCTCTGGGTTCGGCTGCGGCTTCTTTCGGC	540	
Qy	541	GATCGCAGCGCTCTCTTCGGAGTTCCGCGCGCTTCTCTCGACAAATCCCGGGAAC	600	
Db	541	GATCGCAGCGCTCTCTTCGGAGTTCCGCGCGCTTCTCTCGACAAATCCCGGGAAC	600	
Qy	601	CCGCTCTGCGCGAGCGTCGACACAGGTACTGGACCAAAATGTCTTCTCAGTCTCTTTCGCC	660	
Db	601	CCGCTCTGCGCGAGCGTCGACACAGGTACTGGACCAAAATGTCTTCTCAGTCTCTTTCGCC	660	
Qy	661	TGCTCCCTAGCTATTTCGAGAGCTGCTGCTTTGGAGCCAGCTTTCGGGGGCTGATTC	720	
Db	661	TGCTCCCTAGCTATTTCGAGAGCTGCTGCTTTGGAGCCAGCTTTCGGGGGCTGATTC	720	
Qy	721	CTACCAAGTGGCTAGATTTTCCGCAAGTTACAGCGCCCTGAGAGCTTCGAGAGGACGGG	780	
Db	721	CTACCAAGTGGCTAGATTTTCCGCAAGTTACAGCGCCCTGAGAGCTTCGAGAGGACGGG	780	
Qy	781	AGGAATCTGACGCTCCACCGGTTCGAGAAGTCTCTGAGTTACATCGCGGCTGGACTCTTCG	840	
Db	781	AGGAATCTGACGCTCCACCGGTTCGAGAAGTCTCTGAGTTACATCGCGGCTGGACTCTTCG	840	
Qy	841	CGAAGACTCGGCTGCTGAGTACTCTTTGCAATTGGCTAGAGAGGGAATCCAGTGGCAGTGC	900	
Db	841	CGAAGACTCGGCTGCTGAGTACTCTTTGCAATTGGCTAGAGAGGGAATCCAGTGGCAGTGC	900	
Qy	901	TGCTCTCAGACTGGAGTTAACTCAAGCCGAGGAAGAGCTTTAGACTCTGACGCGC	960	
Db	901	TGCTCTCAGACTGGAGTTAACTCAAGCCGAGGAAGAGCTTTAGACTCTGACGCGC	960	
Qy	961	CCACTTTCTCTCTGGAGCAGCAGCTGTGGGGAGTGGAGTTGCTGCCAGTAGCTTCAAG	1020	

Db 961 ||||| CCACTTTCTCTCGAGCAGCAGCTGTGGGAGTGGAGTTGCTGCCAGTAGCCTTCAAG 1020  
Qy 1021 CTGGTCTAGTCTCCACCGAGAACTTGACTCTTCATCTCTGGGCTCTGAGCGTTGAGA 1080  
Db 1021 CTGGTCTAGTCTCCACCGAGAACTTGACTCTTCATCTCTGGGCTCTGAGCGTTGAGA 1080  
Qy 1081 GCTTAGGTAAATTTCAAGGTAGTTTCTTATCTCTGAAACCTTCTTACCTGGAATACCTTC 1140  
Db 1081 GCTTAGGTAAATTTCAAGGTAGTTTCTTATCTCTGAAACCTTCTTACCTGGAATACCTTC 1140  
Qy 1141 CCCAGTTAGGCTGGCTGTCTGAGCAGCGCTGGAGTGGCAGTTTGTGGTTTCGGAA 1200  
Db 1141 CCCAGTTAGGCTGGCTGTCTGAGCAGCGCTGGAGTGGCAGTTTGTGGTTTCGGAA 1200  
Qy 1201 CACTAAACCCAGAGAGCTGTCTTCTTCTGAAGATGGTTGTTCACCTCAGCGCTTGGGG 1260  
Db 1201 CACTAAACCCAGAGAGCTGTCTTCTTCTGAAGATGGTTGTTCACCTCAGCGCTTGGGG 1260  
Qy 1261 CAGAGATGTCGGCAAACCGCTTGGAGAAAGTGTCCGCTCTCTTACAGAAAGGCTGCGG 1320  
Db 1261 CAGAGATGTCGGCAAACCGCTTGGAGAAAGTGTCCGCTCTCTTACAGAAAGGCTGCGG 1320  
Qy 1321 AAATCCACCACCTTCGTATGAAAACGGCTAGAAATTCCTCCAGGCTTAAACAAAGGGCAAGAT 1380  
Db 1321 AAATCCACCACCTTCGTATGAAAACGGCTAGAAATTCCTCCAGGCTTAAACAAAGGGCAAGAT 1380  
Qy 1381 TACCACCCCTGACCAAGATTAATGGCTATCATAGCTTGGAGAGGAAACATAACCTTCTCC 1440  
Db 1381 TACCACCCCTGACCAAGATTAATGGCTATCATAGCTTGGAGAGGAAACATAACCTTCTCC 1440  
Qy 1441 GGATGGACCCACAACTTGCA CAGATAACCCAGACAGGCGGTGTCCCTGTCTGACAGACA 1500  
Db 1441 GGATGGACCCACAACTTGCA CAGATAACCCAGACAGGCGGTGTCCCTGTCTGACAGACA 1500  
Qy 1501 GGCGGGAGCCACTGAGNAAAAACAGAAATTTGGTGAATTCAGAAAGTTTTCAGAGGCCCCC 1560  
Db 1501 GGCGGGAGCCACTGAGNAAAAACAGAAATTTGGTGAATTCAGAAAGTTTTCAGAGGCCCCC 1560  
Qy 1561 AGGGAAGCAGTCTGTTTGTGAAATACCCGTGGAAAAAGAAATGTGAAGAGACCACTA 1620  
Db 1561 AGGGAAGCAGTCTGTTTGTGAAATACCCGTGGAAAAAGAAATGTGAAGAGACCACTA 1620  
Qy 1621 ATGCAACTGACCTCTCAGATAGAGAGAGAGCCCTTCTGTTTCTTACAGACCAAGTTGTA 1680  
Db 1621 ATGCAACTGACCTCTCAGATAGAGAGAGAGCCCTTCTGTTTCTTACAGACCAAGTTGTA 1680  
Qy 1681 GCAACAACTGATAGATTAATTTTGGAGGCGCCCGCAGTGACTTGGAAAGCCAGCTCTG 1740  
Db 1681 GCAACAACTGATAGATTAATTTTGGAGGCGCCCGCAGTGACTTGGAAAGCCAGCTCTG 1740  
Qy 1741 ATTCGAAAGTGAGATTTGGGCGGAGGAACTGTGAGGACGATGGCTTTTGATAGCGATGGCT 1800  
Db 1741 ATTCGAAAGTGAGATTTGGGCGGAGGAACTGTGAGGACGATGGCTTTTGATAGCGATGGCT 1800  
Qy 1801 CCCTGTCTGAAATCAGACGTGGAACAGGACTCGGAAGGCTTTCACCTTTGGAACTCTTTC 1860  
Db 1801 CCCTGTCTGAAATCAGACGTGGAACAGGACTCGGAAGGCTTTCACCTTTGGAACTCTTTC 1860  
Qy 1861 ACAGTGTAGATTCCTTAACAAACCCAAACTTTACGCCACGATTCAGACGCTGCCAGAA 1920  
Db 1861 ACAGTGTAGATTCCTTAACAAACCCAAACTTTACGCCACGATTCAGACGCTGCCAGAA 1920  
Qy 1921 TTGCCCCCAGAGACCCATCAGATTCAGGACATCTGGTCTGGCAGCTGTGGTGTAGGGA 1980  
Db 1921 TTGCCCCCAGAGACCCATCAGATTCAGGACATCTGGTCTGGCAGCTGTGGTGTAGGGA 1980  
Qy 1981 GCTGTGAGGAGACCCCTTCCGAGACCCCGACCATAGTTCCGGGGAGGAAGATGACT 2040  
Db 1981 GCTGTGAGGAGACCCCTTCCGAGACCCCGACCATAGTTCCGGGGAGGAAGATGACT 2040  
Qy 2041 GGGAAACGAGTGCAGATGAAGCAGAGAAATCTTAAATTTGTGGAATCTTTTCTGTCAATCTG 2100

Db 2041 GGGAAACGAGTGCAGATGAAGCAGAGAAATCTTAAATTTGTGGAATCTTTTCTGTCAATCTG 2100  
Qy 2101 AGGACCCCTTCAAAACCTTTTAAATTTTAAAGGCTCTCTTTCAACCGTCAAGGAAGAAATTTGA 2160  
Db 2101 AGGACCCCTTCAAAACCTTTTAAATTTTAAAGGCTCTCTTTTCAACCGTCAAGGAAGAAATTTGA 2160  
Qy 2161 AGGCGCTGAGGACTCAAGGCTCTTCTGAGGTCAAGTGGCCTTCTCTGGCCATCATTA 2220  
Db 2161 AGGCGCTGAGGACTCAAGGCTCTTCTGAGGTCAAGTGGCCTTCTCTGGCCATCATTA 2220  
Qy 2221 CCTTACTTCTTCTGTAAGGCCAGCTGTAGAGCCAAAGAGATAATTTGTCCAGGCTGTG 2280  
Db 2221 CCTTACTTCTTCTGTAAGGCCAGCTGTAGAGCCAAAGAGATAATTTGTCCAGGCTGTG 2280  
Qy 2281 GGCTGGGTGAGGCTCTTGTCTGGAGAAAGATACACCCATATCAAGAGAAAAAAGGTAACTT 2340  
Db 2281 GGCTGGGTGAGGCTCTTGTCTGGAGAAAGATACACCCATATCAAGAGAAAAAAGGTAACTT 2340  
Qy 2341 TCCTGGAAAGATTTACTGAGTATTTATTAAGTGGTGAATGAGGATCGCAAAAGGACCATGGG 2400  
Db 2341 TCCTGGAAAGATTTACTGAGTATTTATTAAGTGGTGAATGAGGATCGCAAAAGGACCATGGG 2400  
Qy 2401 AAGAAATTTGCAAGGGATGGATGAGGTTCCAGAAAAACGAATTTCAAGAAACAGAAAGTTGCCA 2460  
Db 2401 AAGAAATTTGCAAGGGATGGATGAGGTTCCAGAAAAACGAATTTCAAGAAACAGAAAGTTGCCA 2460  
Qy 2461 TTGGCTACTCTTGGCCCTTTGAGCAGACAGAAAAAATGTTTAAATAGACTGAGGATCGAGT 2520  
Db 2461 TTGGCTACTCTTGGCCCTTTGAGCAGACAGAAAAAATGTTTAAATAGACTGAGGATCGAGT 2520  
Qy 2521 CAAAGGACTTACTGTTGTA CAGCAATGTTAAGAGTGAACAGGCTGCACACCCGTGCCAC 2580  
Db 2521 CAAAGGACTTACTGTTGTA CAGCAATGTTAAGAGTGAACAGGCTGCACACCCGTGCCAC 2580  
Qy 2581 TCTGTCTCTTACTTGAAGTTTCCCTTAAACAAACACACACACACACACACACACACACACATG 2640  
Db 2581 TCTGTCTCTTACTTGAAGTTTCCCTTAAACAAACACACACACACACACACACACACACACATG 2640  
Qy 2641 TTTTAAAGAAAAACAACTTGTATCTAGAGATGCAAGTTTGAATTTTGGGTAAATGTGTCT 2700  
Db 2641 TTTTAAAGAAAAACAACTTGTATCTAGAGATGCAAGTTTGAATTTTGGGTAAATGTGTCT 2700  
Qy 2701 CATTAGAAAAACCAACTCCCGATTAATGAAGATCTCTTATCTGTAAATCCTCTCTTTTCCCTA 2760  
Db 2701 CATTAGAAAAACCAACTCCCGATTAATGAAGATCTCTTATCTGTAAATCCTCTCTTTTCCCTA 2760  
Qy 2761 TTTAGTTGGATGTGGGTTTGTCTTTTGGAGGGCTCTCACTGCATAAATTTCTGTTTGGCC 2820  
Db 2761 TTTAGTTGGATGTGGGTTTGTCTTTTGGAGGGCTCTCACTGCATAAATTTCTGTTTGGCC 2820  
Qy 2821 TAGTCTTATGTATGTAGACCAAGTTGGCTTTGAAGTTGGTATTTTCTTGGCTCTGTTTCC 2880  
Db 2821 TAGTCTTATGTATGTAGACCAAGTTGGCTTTGAAGTTGGTATTTTCTTGGCTCTGTTTCC 2880  
Qy 2881 TGTGCCACCAATGCCCAGCTGGAAAGTGTTTTTTAAATATTCCTTCTGCTTACCTGGGGTGAGAG 2940  
Db 2881 TGTGCCACCAATGCCCAGCTGGAAAGTGTTTTTTAAATATTCCTTCTGCTTACCTGGGGTGAGAG 2940  
Qy 2941 TTTTAAATTTTGCATCTTTCACAAACCGTTTCTTCTGGAGGCGAGGGCTGTGCTGTTTACATAC 3000  
Db 2941 TTTTAAATTTTGCATCTTTCACAAACCGTTTCTTCTGGAGGCGAGGGCTGTGCTGTTTACATAC 3000  
Qy 3001 AGGCTTCAGGCCAGCTAGGCTTATGTGGTGAGACCTAGTCTTAGAGGACAAACAGAACAA 3060  
Db 3001 AGGCTTCAGGCCAGCTAGGCTTATGTGGTGAGACCTAGTCTTAGAGGACAAACAGAACAA 3060  
Qy 3061 AACAGTCAGGTTACTGTGGAAAACCTGAGGCAAGGATAGGAAAGTCCAGGTATGCTTGGGA 3120  
Db 3061 AACAGTCAGGTTACTGTGGAAAACCTGAGGCAAGGATAGGAAAGTCCAGGTATGCTTGGGA 3120  
Qy 3121 ACTCAGTGAATCTCTGAATCAAGCCAGATGGGGAGTTAGCAAGACCTTTTACACTC 3180  
Db 3121 ACTCAGTGAATCTCTGAATCAAGCCAGATGGGGAGTTAGCAAGACCTTTTACACTC 3180





Db	5341	CTTTTCAGCTCAGTCATTTTACGTGTAACTGGGAGATTGATAGCTCAGATTATTTGT	5400
Qy	5401	ATATAATTAATTAACCTAATCTGTAATTTGTAATAAATATATTGCAATTTATTAATAAAAAA	5460
Db	5401	ATATAATTAATTAACCTAATCTGTAATTTGTAATAAATATATTGCAATTTATTAATAAAAAA	5460
Qy	5461	AAAAAAAA 5468	
Db	5461	AAAAAAAA 5468	
RESULT 2			
ADQ8457	ID	ADQ8457 standard; cDNA; 2942 BP.	
XX	AC	ADQ8457;	
XX	AC		
DT	07-OCT-2004	(first entry)	
XX		Human GADD34-like (GADD34L) cDNA.	
XX		Oxidative stress; neuronal ischaemia; heart ischaemia; renal damage;	
KW		autoimmune disease; neurodegenerative disorder; therapy; GADD34-like;	
KW		GADD34L; eIF2alpha-specific regulatory subunit; phosphatase; human; gene;	
XX	88.		
XX		Homo sapiens.	
XX			
PH	Key	Location/Qualifiers	
FT	CDS	407..2548	
FT	FT	/*tag= a	
FT	FT	/product= "Human GADD34-like (GADD34L) protein"	
XX			
PN	US2004142345-A1.		
XX			
PD	22-JUL-2004.		
XX			
PF	28-AUG-2003; 2003US-00650482.		
XX			
PR	06-SEP-2002; 2002US-0408679P.		
XX			
PA	(ROND/) RON D.		
PA	(JOUS/) JOUSSE C.		
XX			
PI	Ron D, Jousse C;		
XX			
XX	WPI; 2004-552556/53.		
DR	P-PSDB; ADQ8458.		
DR	GENBANK; AK027650.		
XX			
PT	Screening test substances for preventing or treating disease involving		
PT	oxidative stress, by testing test substances for its ability to inhibit		
PT	activity of GADD34L and identifying test substance that inhibits activity		
PT	of GADD34L.		
XX			
PS	Disclosure; SEQ ID NO 1; 30pp; English.		
XX			
CC	The present invention relates to a method of screening several test		
CC	substances for preventing or treating diseases involving oxidative stress		
CC	such as neuronal ischaemia, heart ischaemia, renal damage induced by		
CC	ischaemia or toxins, autoimmune diseases and neurodegenerative disorders.		
CC	The method involves testing the test substances for its ability to		
CC	inhibit the activity of GADD34-like (GADD34L), also referred to as		
CC	eIF2alpha-specific regulatory subunit of phosphatase, and identifying the		
CC	test substance which inhibits the activity of GADD34L. The present		
CC	sequence is human GADD34L cDNA.		
XX			
SQ	Sequence 2942 BP; 732 A; 739 C; 725 G; 746 T; 0 U; 0 Other;		
Query Match 24.2%; Score 1324; DB 12; Length 2942;			
Best Local Similarity 72.5%; Pred. No. 0;			
Matches 1964; Conservative 0; Mismatches 660; Indels 84; Gaps 16;			

QY 1211 AGAGAGCTGCTATCTTTCTGAAGATGGTTGTCAACCTCAGCGCTTGGCGGCAGAGATGTC 1270  
 Db 1171 AGAGAGCAGCTGCTTGAGAGAGGACCAATGTCATCCCGCAGCGCTGAGTGCAGAACTCAT 1230  
 QY 1271 GGCAACCGCCTGGAGAGAGGTGTCGCTCTCTCTACAGAAAGCCCTGCGGAAATCCACCA 1330  
 Db 1231 TCCGCGCTCGTGGCAGGATGTCACCTCTTTCTACGGAAGGCCCTTACCAGAAATTCACCA 1290  
 QY 1331 CCTTCGTATGAACCGCTAGAAATTCCT--CCAGCTTAACAAGGGCAAGAGTTTACCCAC 1387  
 Db 1291 TCCTTCGCATGAACCGCTGGAATTCCTTCAACAGGCTTAACAGGGGCAAGATTTTACCCAC 1350  
 QY 1388 CCCTGACCAAGATAATGCTATCATAGCTCTGGAGAGGAAATACCTTTCTCCGATGGA 1447  
 Db 1351 CCCTGACCAGGATAATGCTATCACAGCTCTGGAGAGGAAACACAGCCTTCTCCGATGGA 1410  
 QY 1448 CCACAACATTTGCAAGATAACCCGAGCAAGCGGTGTCCTGTCGAGACAGGCC--- 1504  
 Db 1411 TCCAAAACACTGCAGAGATAACCCCAACACAGTTTGTCTGCTGCTGGAGACATTCCTGG 1470  
 QY 1505 -----GGAGCCCACTGAGAAAACACAGAAATGGTGATTCAGAAGTTTCAC--- 1551  
 Db 1471 AAACACCAGGAATCCACTGAGAAAATAAGAAATTTAATCTACAGAGGTTCACATTGC 1530  
 QY 1552 -----AGAGCCCCAGGGAAGCAGTCTGTGTTGTGAATTTACCCGTGGAAGAAGA 1600  
 Db 1531 TTTGGAAGAAGAGAGCCCTCTGAGGGCTGTCCATCTAGTCAGATACCTATGGAAGAAGGA 1590  
 QY 1601 ATGTGAAGAGACCACTAATGMACTGACCTCTCAGATAGAGAGAGACCTTCTGTT 1660  
 Db 1591 GCCTGGAGAGGCCGAATAAGTGTAGTTGATTAATCTATACCTAGAAAGGTGACCTTCCCAT 1650  
 QY 1661 TTCTACAGACAGTGTGTAGCAACAACCTCATAGATTATATTTTGGAGGCGCCCCAG 1720  
 Db 1651 TTCTGCCAGACAGCTTTGTAGTAACAACCTCATAGATTATATTTTGGAGGTGCATCCAG 1710  
 QY 1721 TGACTTGAAGCCAGCTCTGATTTCTGAAAGTGAAGATTTGGGCGAGGAACCTGAGGACGA 1780  
 Db 1711 TGACTTGAACAACAGTTCTGATCCAGAAGGTGAGGATTTGGATGAGGAAGCTGAGGATGA 1770  
 QY 1781 TGGCTTTGATAGAGATGGCTCCCTGTCTGAATCAGACGTGGAACAGGACTCGGAAGGCCT 1840  
 Db 1771 TGGTTTTGATAGTAGTCACTGTCTCAGACTCAGACCTTGAACAAGACCCCTGAAGGGCT 1830  
 QY 1841 TCACCTTTGGAACCTTTTCCACAGTGTAGATCCTTTACAAACCCCAAACTTTACAGCCAC 1900  
 Db 1831 TCACCTTTGGAACCTTTTCTCAGTGTAGATCCTTTATANTCCCAAGAACTTTACAGAAC 1890  
 QY 1901 GATTTCAGACGCTGTCAGAAATGCCCCCAGAGACCCCATCAGATTTCAGGACATCTCTGTC 1960  
 Db 1891 AATTTCAGACTGCTGCCAGAAATGTTCTCTGAAGAGCCTTCTGATTCAGAGAAGGATTTGTC 1950  
 QY 1961 TGGCAGCTGTGGTGTAGGGA---GCTGTTCAGAGAGGAGCCCTTCCGGAGACCCCGACCA 2017  
 Db 1951 TGGCAAGTCTGATCTAGAGAAATTCCTCCAGTCTGGAAGCCCTTCTGAGACCCCTGAGCA 2010  
 QY 2018 TAGTTCCGGGGAGAGAGATGATCGGAACCCAGTGCAGATCAAGCAGAGATCTTAAATT 2077  
 Db 2011 TAGTTCTGGGAGAGAGAGATGATCGGAATCTAGTGCAGATGAAGCAGAGATCTCAAACT 2070  
 QY 2078 GTGGAACCTCTTTCTGTCATTTCTGAGGACCCCTACAAACCTTTTAAATTTTAAAGGCTCCTTT 2137  
 Db 2071 GTGGAACCTCATCTGTAATCTGATGACCCCTACAAACCTTTTAAATTTTAAAGGCTCCTTT 2130  
 QY 2138 TCAACCGTTCAGGGAAGAAATGGAAGGCCGTCTAGCACTCAAAAGGCCTTTCTGAGGTGAC 2197  
 Db 2131 TCAAAACATCAGGGGAAAATGGAAGGCTGCTGACTCAAAAGACCCCATCTGAGTCCAT 2190  
 QY 2198 AGTGCCCTTCTCGCCCATCTACTTCTTGTGAAGGCCAGCTGTTTGAAGAGGCA 2257  
 Db 2191 TGTGGCCATTTCTGAGGTGACACCTTTACTTTCTTGTGAAGGTGACGCTGTTTGGGAGGCA 2250  
 QY 2258 AGAAGATAAATGTTCCAGGCTGTGGGCTGGGTGAGGCTCTTGTCTGGAGAAAGATACACCCA 2317

Db 2251 AGAAAGTGAATGTCAGACTCGGTACAGCGTGCAGTCTTTCTCGAGGAACACACACA 2310  
 QY 2318 TATCAAGAGAAAAGTAACTTCTCGGAAGAGTTACTGAGTATTATATTAAGTGGTGA 2377  
 Db 2311 TGTCAAAAGAAAAGGTAACTTCTTGAAGAGTTTACTGAGTATTATATAAGTGGTGA 2370  
 QY 2378 TGAGGATCGCAAGGACCATGGGAAGAAATTTGCAAGGGATGGATGCGAGTTTCCAGAAACG 2437  
 Db 2371 TGAGGATCGCAAGGACCATGGGAAGAAATTTGCAAGGGATGGATGCGAGTTTCCAGAAACG 2430  
 QY 2438 AATTCAAGAAAACAGAAAGTTGCCATTGGCTACTGCTTGGCCCTTTGAGCAGAGAAAAAAT 2497  
 Db 2431 AATTCAAGAAAACAGAAAGTGTATTTGGATATTTGCTTGCATTTTGAACACAGAGAAAGAA 2490  
 QY 2498 GTTTAATAGACT---GAGGATCGAGTCAAGGACCTTACTGTTGCTACAGCAATCTTAAGA 2553  
 Db 2491 GTTTAATAGACTCCAGGAAACATGCTTTCAAGGACCTTAATGTTCTCAAGCAATGTTGAGT 2550  
 QY 2554 AGTGA-ACAGCTGCACACCCGTCGCCACTCTGTCTCTTACTTTCAGA-GTTTCCCTTAAAA 2611  
 Db 2551 TGGCAGCTCTAGTCTCTAGCTAGCATACACTACCTCTTACTGAGAGGTGTCTTTTAAAA 2610  
 QY 2612 ACAAACACTGGCAGCTGTCTTGGACATGTTTTTAAAGAAACAACTTGTATCTAGAGATG 2671  
 Db 2611 ACAAATCTTGGCAGCTGTCTTGCACATTTTTTTTTTTAGAGGAAATGTAACCTTGGATCT 2670  
 QY 2672 CAGTTGATTTATTTTGGGTAATGTGTCTCATAGAA-----ACACCAACT 2717  
 Db 2671 AGTTTAAATTTTTTTTTTTCACAAATATCCCACTCAGAAACATTCAGGTTTGAAGCCAGCC 2730  
 QY 2718 CCGAATATGAAGAAAT-----CTCTATCTGTAATCCTCTCTTTTCTCTATTTAGTTGGAT 2771  
 Db 2731 CTGATATGAAGATGAAGTGAAGTGTGATTTCTAATCCTCCCTTTTGTGATTTAGTTGGAT 2790  
 QY 2772 GTGGGTTT 2779  
 Db 2791 GTGCTTTT 2798

RESULT 3  
 AAH18699  
 ID AAH18699 standard; cDNA; 2942 BP.  
 XX  
 AC AAH18699;  
 XX  
 DT 26-JUN-2001 (first entry)  
 XX  
 DE Human cDNA sequence SEQ ID NO:18964.  
 XX  
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 FN EF1074617-A2.  
 XX  
 PD 07-FEB-2001.  
 XX  
 PF 28-JUL-2000; 2000EP-00116126.  
 XX  
 PR 29-JUL-1999; 99JP-00248036.  
 PR 27-AUG-1999; 99JP-00300253.  
 PR 11-JAN-2000; 2000JP-00118776.  
 PR 02-MAY-2000; 2000JP-00183767.  
 PR 09-JUN-2000; 2000JP-00241899.  
 XX  
 PA (HELI-) HELIX RES INST.  
 XX  
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
 XX WPI; 2001-318749/34.  
 DR  
 XX

PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.

XX Claim 8; SEQ ID NO 18964; 2537pp + Sequence Listing; English.

XX The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention

XX Sequence 2942 BP; 731 A; 741 C; 726 G; 744 T; 0 U; 0 Other;

Query Match 24.2%; Score 1320.8; DB 4; Length 2942;

Best Local Similarity 72.5%; Pred. No. 0;

Matches 1962; Conservative 0; Mismatches 662; Indels 84; Gaps 16;

Qy 151 CCGTCGCGCGCGCGAGGAGGCTCTCTATGTTGGAGCGATCTCACACGCCCTAGG 210  
 Db 96 CCATCGCGAGCGCGTGGGGGACTCTCTATCCATGGTGTGAGCGTGGAGCGGACTAGG 155

Qy 211 ACGTCTCTTCCCTAGCGGGATGGACTAACCGCGGTGCGACCGCTCGCGGGGCTC 270  
 Db 156 GAACCTCTTCTCC -CGCAGGATGGAGTGGATCGATCGCGCGCTATTGCGGGGCTGT 214

Qy 271 TGGGCGGTGCGGTGCAGTCTGTTGCGAAGCGCGCTCTCTGCGGCTCTCTGCGG 330  
 Db 215 TCTTCCCTGTGTTCTGCGCGCGCTGCGCATTCGCTGCGCTCTGTGCGTCTTCTGCTG 274

Qy 331 CGCGGGAATCGGACTGAGTACCCACTCCGTTGGTGGGCAAGCGGAGACTGTGTAGACC 390  
 Db 275 CTCGAAGATCGGCTGGAGCAGCAGCAGCCGCTGGGCAAGGCGGAGACTCTGTAGGCT 334

Qy 391 TCGGATCCAGCTCGGCTGACCGCGCTGAGTCTGTCTCTCTCTGTCTGAGAAGCGGCC 450  
 Db 335 TCCTCCGAATCCCGTCGACCTCCAGCGCGCTGAGCGCGCGCGCGCTTACCTGAGAGCTGTC 394

Qy 451 AAG-GAAGAGAGATGGAGACAGGAACGACAGGCGCCCGAGCGGCTCGGCTCGGCTG 509  
 Db 395 AAGAAAAGAGATGGAGCGGAGCAGGCGGATCGCGAAGACGCGCTTGGCCCTCGGGCG 454

Qy 510 GGCTCTCGGTTCCGGCTGCCCC---TTCTTTCGGCGATCGCAGCGCTCTCTTCGGAGTTC 566  
 Db 455 GGCTTCCGGTTCTGGCCACCCCTTTTCTTCCCTGGCGATCGCAAGCAGGCTCTTCTAAGTTC 514

Qy 567 CCGCGCGCTTCTCTGACAAATCCCGGGAACCTC-----GCTCTGCC 611  
 Db 515 CCGACGCTCTTGGCCCGGAAATCTCCGGGAACCCCACTGCTTCTCTGCGCCAGCCCC 574

Qy 612 GAGCGTGGAGCCAGTACTGACCAAAATGCTTTCTCAGCTCTTTCGCTCTCCCTAGC 671  
 Db 575 GAGACTCGGTCAGTTACTGACGAAACTGCTCTCCAGCTCTTTCGCGGCTCCCGGA 634

Qy 672 CTATTCCAGAGCTGCTGCTTTGGAGCGAGCTTTTCGGGGGCGGCTGATTCCTACAGATGG 731

Db 635 TTGCTTTAGAGAGTGTCTAAATTTGGAGCCAACTTTTGGTGGAAATGTTTCCGACAGATGG 694  
 Qy 732 CTAGATTTTGGCCCAAGATTACAGCGCCCTGAGAGCTTTCGAGAGGAGGAGGAAATCTTGAC 791  
 Db 695 CTAGATTTTGGTGGATCTACAGCGCCCTGAGAGCCCTGAGAGGACGGGAGAAACCAAGCC 754

Qy 792 GCTCCCAACGGTGCAGAAAGTCTCTGAGTTACACTTCGCGCTGGACTTTCGCGAAGACTCGC 851  
 Db 755 GCGCCACAGCGCAGAAATCTTTGAGTT-CGCTGCACTCGACTCTCTC---AGACCCCTC 810

Qy 852 GTCGTCAGTACTCTTGCATTTGGTAGAGGAGGACTCCAGTGCAGTGTCTGCTCTCAG 911  
 Db 811 GGTCAACAGCTCCCTTTGATTTGGCTAGAGGAAGGATCCACTGGCAATACTTCGCCCCAGA 870

Qy 912 -CTGGAAGTTAAACTCAAGGCCCAGGAAGAGCTTTTAGACTCTGCAAGCGCCCACTTTCT 970  
 Db 871 CCTAAATTTGGAGCTTTAGGCCCAAGGAGTGTCTTTGGACCTTGCAGCAGCAGGCTTTCT 930

Qy 971 CTTGGAGCAGCAGCTGTGGGGAGTGGAGTTGTCTGCCAGTAGCTTTCAAGCTGTCTAGT 1030  
 Db 931 CTTAGAGCAGCAGCTGTGGGGAGTGGAGTGTGTGCCAGTAGCTTTCAATCCCGTCTGTA 990

Qy 1031 CTCCACCGAGAACTTGACTTCTTCATCTCTGCGGCTCTGAGGCTTTCAGAGCTTAGGTAA 1090  
 Db 991 CTCTAACCGGGAATCTTGGCTCTTTCGCGCTCTTGGGCTCTTAAACATTTCAAGCAGTAGACGA 1050

Qy 1091 TTTTCAAGGTAGTTTCTCTCTCTGAAACCTTCTCTACCTTGGACTACTTCTCCCGAGTTAG 1150  
 Db 1051 TTTTCAAGGTAGTTTCTCTCTCTGAAACCTTCTCTACCTTGGACTACTTCTCCAGGCTAGA 1110

Qy 1151 GCTGCGCTGTGAGAGCAGCGCTGGAGGTGGCCAGTTTGTGGTTCCTCGAACACTAACCC 1210  
 Db 1111 AGTCAGCTATCAGAACAGTGTAGGAAATAGCGAGGTAGTGGCTTCCAGACACTAACCC 1170

Qy 1211 AGAGAGTGTCTATCTTTCTGAGAGTGTGTCTACCTCTGAGCGCTTTCGCGGAGAGATGTC 1270  
 Db 1171 AGAGAGCAGCTGTCTGAGAGAGGACCAATTTGTCTATCCCGCGCTGAGTGCAGAACTCAT 1230

Qy 1271 GGCACCGCTGTGAGAGAGTGTCCGCTCTCTCTACAGAGGCTGCGCGAATTCACCA 1330  
 Db 1231 TCCGCTCTGTGAGAGAGTGTCCAGCTCTTTCTAGGAAGGCTTACAGAAATTCACCA 1290

Qy 1331 CTTCTGTATGAACCGCTAGAAATTCCT---CCAGGCTAACAAAGGCAAGAGTTACCCAC 1387  
 Db 1291 TCTTCGATGAACCGCTGGAATTCCTTCAACAGGCTAGCAAGGCGCAAGATTTACCCAC 1350

Qy 1388 CCTGACCAAGATTAATGGCTATCATAGCTTGGAGGAGGAAACATAAATCTTCTCCGATGGA 1447  
 Db 1351 CCTGACCAAGATTAATGGCTATCCACAGCTTGGAGGAGGAAACACAGCTTCTCCGATGGA 1410

Qy 1448 CCACAACTTGCACAGATAACCCAGCAGCGGCTGCTCCCTGCTGCGACAGAGCC--- 1504  
 Db 1411 TCCAAACACTGCAGAGATAACCCAAACAGATTTGTCTGCTGCTGGAGACATTCCTGG 1470

Qy 1505 -----GGAGCCCACTGAGAAAAAACCAGAAATTTGGTGAATTCAGAAAGTTTCAC--- 1551  
 Db 1471 AAACACCCAGGAATCCACTGAAGAAAAATAGAAATTTAATTAACCTACAGAGGTTCCACTTGC 1530

Qy 1552 -----AGAGCCCCCAGGGAAGCAGTCTGTGTTTGTGAATTTACCGTGGAAAAAGA 1600  
 Db 1531 TTTTGAAGAGAGAGAGCCCTTCTGAGGCTGTCTCATCTAGTGAATACCTATGGAAGGA 1590

Qy 1601 ATGTGAAGAGGACCACTAAATGCAACTGACCTCTCAGATAGAGGAGAGAGCCCTTCTGT 1660  
 Db 1591 GCTGAGAGGCGCGAATAGTGTAGTTAGTTACTCATCTACCTAGAGGAGTACCTTCCAT 1650

Qy 1661 TTCTACAGACCAAGTTTGTAGCAACAACTGATAGATTAATTTTGGAGGCGCCCCAG 1720  
 Db 1651 TTCTGCCAGACAGCTTGTAGTAACAACTGATAGATTAATTTTGGAGGTCATCCAG 1710

Qy 1721 TGACTTGGAGAGCCAGCTCTGATTTCTGAAAGTGAAGATTGGGGCGAGAACTGTAGAGACGA 1780

Db 1711 TGACCTGGAAACAAGTTCTGATCCAGAAAGGTGAGGATTTGGGATGAGGAAGCTGAGGATGA 1770  
Qy TGGCTTTTGTATAGCACTGGCTCCTCTCTGCTGAATCAGACGTGGAACAGGACTCGGAAGGCTT 1840  
Db TTTTGTATAGTATAGTACTCTGCTGAGACTCAGACCTTGAAACAGACCTCTGAAGGCTT 1830  
Qy TCACCTTTGGAACTCTTTTCCACAGGTGATGATCCCTTACAAACCCCAAACTTTTACAGCCAC 1900  
Db TACCTTTTGGAACTCTTTTCTGAGTGTAGATCTTTATTAATCCCAAGAACTTTTACAGCAAC 1890  
Qy GATTTCAGACGGCTGCCAAGTTGCCCCAGAGACCCATCAGATTCAGGGACATCCTGGTC 1960  
Db AATTTCAGACTGTGTCAGAAATTTGTCCTGAAGAGCTTTCTGATTCAGAGAAGGATTTGTC 1950  
Qy TGGCAGCTGTGTGTAGGGA---GCTGTCCAGAGGGAACCCCTTCGGAGACCCCGACCA 2017  
Db TGGCAAGTCTGATCTAGAGAATTCCTCCAGTCTGGAAGCCTTCTTGAGACCCCTGAGCA 2010  
Qy TAGTTCCGGGAGGAAGATGACTCTGGAAACCGAGTGCAGATGAAGCAGAGAATCTTAAATT 2077  
Db TAGTTCTGGGAGGAAGATGACTCTGGAAATCTAGTGCAGATGAAGCAGAGAGTCTCAAACT 2070  
Qy GTGGAATCTTTCTGTCTCATCTTGAGGACCCCTTACAACCTTTTAAATTTTAAAGGCTCCTTT 2137  
Db GTGGAATCTTTCTGTCTCATCTTGAGGACCCCTTACAACCTTTTAAATTTTAAAGGCTCCTTT 2130  
Qy TCAACCGTCAGGGAAGAAATGGAAGGCGCTCAGGACTCAAAAGGCTCTTCTGAGGTCAAC 2197  
Db TCAACATCAGGGGAANAATGAGAAGGCTGTCTGACTCAAAAGACCCCATCTGAGTCCAT 2190  
Qy AGTGCCCTTCTGTGCCCATCATACCTTACTTTCTGTAAGGCCAGCCAGCTGTGTAGAGACCA 2257  
Db TGTGGCCATTTCTGAGTGTCAACCTTACTTTCTTGTAAAGTGCAGCTGTGTGGGGAGCCA 2250  
Qy AGAGATAATTTCTCCAGGCTGTGGCTGGGTGAGGCTCTTCTGTCGAGAAAGATCACCCCA 2317  
Db AGAAAGTGAAATGTCAGACTCGTACAGCGTGACGTTCTTTCTGAGGAAGACACACACA 2310  
Qy TATCAAGAGAAAAAGGTAACTTCTCTGGAAGAGTTACTGAGTATATATAAGTGTGTA 2377  
Db TGTCAAAAGAAAAAGGTAACTTCTCTGGAAGAGTTACTGAGTATATATAAGTGTGTA 2370  
Qy TGAGATCGCAAAAGCAACATGGAAGAAATTTGCAAGGATGGATGCGAGTTCCAGAAACG 2437  
Db TGAGATCGCAAAAGCAACATGGAAGAAATTTGCAAGGATGGATGCGAGTTCCAGAAACG 2430  
Qy AATTCAAGAACAGAAAGTTGCAATTTGGCTACTGCTTGGCTTTGAGCAGCAGACAGAAAAAT 2497  
Db AATTCAAGAACAGAAAGTTGCAATTTGGATATGCTTTGACATTTGAACACAGAGAAGAAAT 2490  
Qy GTTTAATAGACT---GAGGATCGAGTCAAAAGGACTTACTGTTGTACAGCAATGTTAAGA 2553  
Db GTTTAATAGACTCCAGGGAACATGCTTCAAAAGGACTTAATGTTCTCAAGCAATGTTGAGT 2550  
Qy AGTGA-ACAGCTGCAACCGTGCCACTCTGCTCTTACTTGTAGA-GTTTCCCTTAAAA 2611  
Db TGGCAGCCTGTAGTCTCTAGCTAGCATACACTACCTCTTACCTGAGAGTGCTTTTAAAA 2610  
Qy ACAACACTGGCAGCTGCTTGGACATGTTTAAAGAAACAACCTTGATCTAGAGATG 2671  
Db ACAAACTTGGCAGCTGCTTGGACATGTTTAAAGAAACAACCTTGATCTAGAGATG 2670  
Qy CAGTTTGAATTTTGTGGTAAATGTGCTCATTTAGAA-----ACACCAACT 2717  
Db AGTTTAAATTTTGTGGTAAATGTGCTCATTTAGAAATTTTGTAGAGAAATGTAACCTTGGATCT 2730  
Qy CGATAATGAAGAAAT-----CTCTTATCTGTAACTCTCTTTTCTTATTTAGTTGAT 2771  
Db CTGATAATGAAGGATGAACCTAGTGTGATTTCTTAATCTCTCCCTTTTGTGATTTAGTTGAT 2790  
Qy GTGGGTTT 2779  
Db GTGCTTTT 2798

RESULT 4

ID AAC98158 standard; cDNA; 2324 BP.  
XX AAC98158;  
AC AAC98158;  
DT 09-MAR-2001 (first entry)  
XX Human colon cancer antigen nucleotide sequence SEQ ID NO:168.  
DE Human; colon cancer; colon cancer antigen; diagnosis; detection;  
XX identification; cytostatic; cardioactive; neuroprotective; vulnery;  
KW immunomodulatory; muscular; gynaecological; gastrointestinal;  
KW nephrotropic; antinfecive; antibacterial; gene therapy; wound;  
KW neural disorder; immune system disorder; muscular disorder;  
KW reproductive disorder; gastrointestinal disorder; renal disorder;  
KW infectious disease; cardiovascular disorder; ss.  
XX Homo sapiens.  
XX W0200055351-A1.  
XX 21-SEP-2000.  
XX 08-MAR-2000; 2000WO-US005883.  
XX 12-MAR-1999; 99US-0124270P.  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX Rosen CA, Ruben SM;  
XX WPI: 2000-587534/55.  
XX P-PSDB; AAB53401.  
XX Colon cancer associated gene sequences, referred to as colon cancer  
XX antigens, useful for the treatment, prevention, and diagnosis of colon  
XX disorders such as colon cancer.  
XX Claim 1; Page 595; 2104pp; English.  
XX AAC97991 to AAC98763 encode the human colon cancer associated proteins,  
XX called human colon cancer antigens, given in AAB53234 to AAB54006. The  
XX human colon cancer antigens can have cytostatic, cardioactive, muscular;  
XX neuroprotective, immunomodulatory, gynaecological, gastrointestinal,  
XX vulnery, nephrotropic, antinfecive and antibacterial activities, and  
XX can be used in gene therapy. The colon cancer antigen polynucleotides,  
XX proteins and antibodies to the proteins are useful for the prevention,  
XX treatment and diagnosis of colon disorders, such as colon cancer. The  
XX polynucleotides may be used in diagnostics and research, such as for  
XX chromosome identification, and as hybridisation probes. The proteins may  
XX also be used to prevent diseases such as neural disorders, immune system  
XX disorders, muscular disorders, reproductive disorders, gastrointestinal  
XX disorders, wounds, renal disorders, infectious diseases, and  
XX cardiovascular disorders. AAC98764 to AAC98772 and AAB54007 represent  
XX sequences used in the exemplification of the present invention  
XX Sequence 2324 BP; 615 A; 566 C; 562 G; 574 T; 0 U; 7 Other;

Query Match 21.7%; Score 1189.2; DB 3; Length 2324;  
Best Local Similarity 75.3%; Pred. No. 1.2e-295;  
Matches 1642; Conservative 4; Mismatches 477; Indels 59; Gaps 11;

Qy 519 TTCCGGCTGCCCTTCCCTTCGGCGATCGACGCTGCTTCGGAGTTCCCGCGCTTCC 578  
Db 43 TTTCAACCTCTGTAGTCGGCGATCGAAGCAGGCTCTTCTAAGTTCCCGACGCTCTT 102  
Qy 579 TCTCGACAAAATCCCGGAACTCCGCTCT-----GCCGAGCGGTGGAAC 623  
Db 103 GGCCCGGAAAACCTCGGGAAACCCACACTGTTTCTCTGCCCGCCGAGACTCGGGTC 162



XX	16-OCT-2001	(first entry)		QY	1719	AGTGACTTGGAGCCAGCTCTGATTTCTGAAGTGGAGGATTTGGGGCGAGGAACTCTGAGGAC	1778
XX	DE			Db	1	AGTGACCTGGAAAACAAGTCTGATCCAGAAAGTGGAGATTTGGGATGAGGAAGCTGAGGAT	60
XX	KW	Human protein encoding cDNA sequence SEQ ID NO:224.		QY	1779	GATGGCTTTGATAGCGATGGCTCCCTGCTCTGATATCAGACCTGGAAAGGACTCGGAAGGC	1838
XX	KW	Human; cancer; ulcer; HIV infection; human immunodeficiency virus; anti-inflammatory; antirheumatic; antiarthritic; immunosuppressive; antibacterial; endocrine; cardiant; central nervous system; virucide; anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; anaemia; antiagregant; haemostatic; vulnery; antitumor; osteopathic; eczema; dermatological; antiallergic; antiasthmatic; antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic; antiparkinsonian; infection; immunostimulant; gene therapy; antisense therapy; vaccine; inflammation; antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis; cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity; genetic disease; haematopoietic disorder; platelet disorder; asthma; thrombocytopenia; osteoporosis; severe combined immunodeficiency; allergic rhinitis; diabetes; multiple sclerosis; depression; Alzheimer's disease; Parkinson's disease; neurodegenerative disorder; neurological disorder; ss.	QY	1779	GATGGCTTTGATAGCGATGGCTCCCTGCTCTGATATCAGACCTGGAAAGGACTCGGAAGGC	1838	
XX	KW			Db	61	GATGGTCTTGTATAGTATGATCTCAGCTCAGACCTTGAACCAAGACCTCTGAAGGG	120
XX	KW			QY	1839	CTTCACCTTTTGGAACTCTTCCACAGTGTAGATCTTACAAACCCCAAACTTTTACAGCC	1898
XX	KW			Db	121	CTTCACCTTTTGGAACTCTTCTGCAGTGTAGATCTTATTAATCCCAAGACTTTTACAGCA	180
XX	KW			QY	1899	ACGATTCAGACGGCTGCCAATTTGCCCCAGAGACCCATCAGATCAGAGACATCTCTGG	1958
XX	KW			Db	181	ACAAATTCAGACTGCTGCCAGAAATTTGTTCTCTGAAGAGCTTCTGATTCAGAGAGGATTTG	240
XX	KW			QY	1959	TCGTGGACAGCTGTGGTGTAGGGA---GCTGTTCAGGAGGACCCCTTCCGGAGACCCCGAC	2015
XX	KW			Db	241	TCTGGCAAGTCTGATCTAGAGAAATTCCTCCAGTCTCGAAGCCCTTCTCGAGACCCCTGAG	300
XX	KW			QY	2016	CATAGTTCCGGGAGGAGGATGACTGGGAAACCGAGTGCAGATGAAGCAGAGAAATCTTAAA	2075
XX	OS	Homo sapiens.		Db	301	CATAGTTCTGGGAGGAGGAGTACTGGGAATCTAGTGCAGATGAAGCAGAGAGCTTCTCAA	360
XX	PN	WO200153455-A2.		QY	2076	---TTGTGGAACCTCTTCTGTCTCA---TTCTGAGGACCCCTTACAACTTTTAAATTTAAG	2129
XX	PD	26-JUL-2001.		Db	361	ACTGTGGAACTTCAATCTGTAAATCTGGATGGACCCCTTACAACTTTTAAATTTAAG	420
XX	PF	22-DEC-2000; 2000WO-US035017.		QY	2130	GCTCTCTTTCACCGTTCAGGAGGAGGATTTGGAAGGCGCTCAGGACTCAAGAGGCTCTTCT	2189
XX	PR	23-DEC-1999; 99US-00471275.		Db	421	GCTCTCTTTCACCAACATCAGGGGAAAATGAGAAAGGCTGTCTGAGACTCAAAAGACCCCATCT	480
XX	PR	21-JAN-2000; 2000US-00488725.		QY	2190	GAGGTTCACAGTGGCTTCTCTGGCCATCATACCTTACTTCTTGTAGAGGCCAGCTGTTA	2249
XX	PA	(HYSE-) HYSEQ INC.		Db	481	GAGTTCACAGTGGCTTCTCTGGCCATCATACCTTACTTCTTGTAGAGGCCAGCTGTTG	540
XX	PI	Tang YT, Liu C, Drmanac RT;		QY	2250	GAGAGCCAAAGAAATAATTTGTCAGGCTGTGGGCTGGGTGAGGCTCTTGTCTGGAGAAAGA	2309
XX	DR	WPI; 2001-457603/49.		Db	541	GGGAGCCAAAGAAATGAAATGTCAGACTCGGTACAGCGTGCAGCTTCTTCTCTGGAGAAAGA	600
XX	DR	P-PSDB; AAM25448.		QY	2310	TACACCATATCAGAGBAAAAGGTAACCTTCTCTGGAAGAGTTCAGTGTAGTATTATATA	2369
XX	PT	Isolated human polynucleotides encoding polypeptides, useful for the treatment and diagnosis of e.g. cancer, ulcers and HIV infection.		Db	601	CACACATGTCAAGAGAAAGGTAACCTTCTCTTGAAGAGGTTACTGTAGTATTATATA	660
XX	PS	Claim 1; Page 398; 1217pp; English.		QY	2370	AGTGGTGTATGAGGATCGCAAAGGACCATGGGAAGAATTTGCAAGGGATGGATGCAAGGTTT	2429
XX	CC	AAH99166 to AAH99904 encode the human proteins given in AAM25225 to AAM25963. The proteins can have activities based on the tissues and cells they are expressed in, such as: anti-inflammatory; antirheumatic; antiarthritic; immunosuppressive; antibacterial; endocrine; cardiant; central nervous system; virucide; anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; antiagregant; haemostatic; vulnery; antitumor; osteopathic; dermatological; antiallergic; antiasthmatic; antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic; antiparkinsonian; and immunostimulant. The proteins and polynucleotides encoding them can be used in gene therapy, antisense therapy and vaccine production. The proteins and polynucleotides are useful for screening for agonists or antagonists of a protein and for the treatment and diagnosis of disorders associated with the activity of a protein e.g. inflammation, rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction, neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal infections, autoimmunity, genetic diseases, haematopoietic disorders, anaemia, platelet disorders, thrombocytopenia, wounds, burns, ulcers, osteoporosis, severe combined immunodeficiency, eczema, allergic rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression, Alzheimer's disease, Parkinson's disease, neurodegenerative and neurological disorders	QY	2490	GAAAGATGTTTAAATAGACTCT---GAGGATCGAGTCAAGAGGACTTACTGTGTACAGCAA	2545	
XX	CC			Db	781	GAAAGATGTTTAAATAGACTCTCAGGGAACATGCTTCAAGAGACTTAATGTCTCAAGCAA	840
XX	QY	Sequence 883 BP; 263 A; 173 C; 218 G; 229 T; 0 U; 0 Other;		QY	2546	TGTTAAGAGTGAAC 2560	
XX	QY			Db	841	TGTTGAGTTGGCAGC 855	
XX	RESULT 6			XX	AAH06623		
XX	ID	AAH06623 standard; cdna; 772 BP.		XX	AC	AAH06623;	
XX	XX			XX	DT	26-JUN-2001 (first entry)	
XX	XX	Human cdna clone (5'-primer) SEQ ID NO:3458.		XX	DE	Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.	
XX	XX			XX	KW		
XX	XX			XX	XX		

Query Match 9.9%; Score 542.6; DB 4; Length 883;  
 Best Local Similarity 80.5%; Pred. No. 5,8e-129;  
 Matches 688; Conservative 0; Mismatches 15; Indels 13; Gaps 4;





PT tissues or cells.  
XX Claim 1; SEQ ID NO 1296; 239pp; English.  
XX  
CC The invention relates to methods for predicting toxic effects of  
CC compounds or the progression of these toxic effects by determining the  
CC global changes in gene expression in tissues or cells exposed to the  
CC toxin and comparing these to gene expression in unexposed tissues or  
CC cells. Also included are methods of predicting at least one toxic effect  
CC of a compound or progression of a toxic effect, preferably the  
CC hepatotoxicity of a compound, comprising detecting the level of  
CC expression in a tissue or cell sample exposed to the compound of two or  
CC more genes listed in the specification, where differential expression of  
CC the genes is indicative of at least one toxic effect or progression. The  
CC method can also be used to identify an agent which modulates the toxic  
CC response and predict cellular pathways that a compound modulates in a  
CC cell. The methods utilize a set of at least two probes (on a solid  
CC support in kit form), where each of the probes comprises a sequence that  
CC specifically hybridizes to a gene listed in the specification, a computer  
CC system comprising a database containing information identifying the  
CC expression level in a tissue or cell sample exposed to a hepatotoxin of a  
CC set of genes comprising at least two genes listed in the specification,  
CC and a user interface to view the information used to present information  
CC identifying the expression level in a tissue or cell of at least one gene  
CC listed in the specification. The method is useful for elucidating global  
CC changes in gene expression and for identifying toxicity markers in  
CC tissues or cell exposed to a known toxin. The genes may be used as  
CC toxicity markers in drug screening and toxicity assays. The genes and  
CC gene expression information may be used as diagnostic markers for the  
CC prediction or identification of the physiological state of tissue or cell  
CC sample that has been exposed to a compound or agent. Hepatotoxicity is  
CC characterized by centrilobular necrosis and steatosis. The present  
CC sequence is an expressed sequence tag (EST) or cDNA derived from a gene  
CC which is differentially expressed in response to a hepatotoxic agent

XX  
SQ Sequence 540 BP; 162 A; 119 C; 85 G; 174 T; 0 U; 0 Other;

Query Match 3.3%; Score 181.6; DB 6; Length 540;  
Best Local Similarity 75.8%; Pred. No. 6.8e-36;  
Matches 388; Conservative 0; Mismatches 69; Indels 55; Gaps 11;  
QY 4960 ATATTATAAAATTTTCAGATTACCTTCACTATCAAAATGAGTAATGAC----- 5011  
DB 499 ATGTTGAAATTTGTTCCAGATTATACCTTCACTATCAAAATGAGTAATGAC 440  
QY 5012 ----TCATGCTGCAGACATGCTCTAGTGGTGGCAGACAGAGGATCT-----TTG 5060  
DB 439 ATAATCAGCCTGCGAGATGCTCAATGGTGGCAAGACAGAGGGTCTAACTGAAGG 380  
QY 5061 ACTGAAGCAGAAAACTGTCTATGTCATCCAGCCCCCAGGAAAGAACACCTCCAAAGGCA 5120  
DB 379 AGAAAAACAAAATTAATCTGTCATGTCAT-CGNAATCCCAAGGNAAGAACACCTCCGAGGCA 321  
QY 5121 GCGAGCAGCAGGCGAGGATGGTGGTCTAGTTGAATACATTC-AAGTCTTGCAATG 5179  
DB 320 -----AGGCAGGCGATGGCGGTCTAGCTGAACACTTACACTCAGTCTTGCAATG 272  
QY 5180 GTGCTTT--AGATCTGTAGCATGTGAGGCTCTGTACAGTGGGG-CCACACTTCTGA 5236  
DB 271 GTGCTTTAGAGATCTCTGTAGCATGTGAGGCTCTGTACAGTGGGGACCCACACTTCCGA 212  
QY 5237 GGGCTGGAATGTGGCAACCTTTATCTAACTTGAATCAAAACCGTCAAAATTTTATTTTT 5296  
DB 211 GTGCTGGAATGTGGCAACCC-TTATCTAACTTGAATCAAAACCGTCAAGATTTATTTTT 153  
QY 5297 TATAATTTAAGAA-----AGATGTGGGGAATGACATTTTTTGAATGGCTTTTTCAGCTCA 5352  
DB 152 TATAATTTAAGAAAAAGTAGACGCGAATGAGAAATTTTTTTTTTTTGAATGAGATGTT 93  
QY 5353 G-----TCATTTTACGTGTAAAC--GTGAGATTTTGAATGCTCAGATTTATATT 5397  
DB 92 GAGTTGGTTTCAGATCATTTTACATGTAACTAGTGTAGATTTTATAGCGGAGATTATATT 33

QY 5398 TGTATATAAATTTATTAACATAATCTGTAATAATTGT 5429  
DB 32 TGTATATAAATTTATTAACATAATCTGTAATAATTGT 1  
RESULT 8  
ID ADB57750/C  
XX ADB57750 standard; DNA; 540 BP.  
XX ADB57750;  
XX 04-DEC-2003 (first entry)  
XX Toxicity-related gene, SEQ ID 2776.  
XX Toxic; toxin; gene expression profile; hepatotoxicity; liver;  
XX drug screening; toxicity assay; ds.  
XX Unidentified.  
XX WO2003064624-A2.  
XX 07-AUG-2003.  
XX 31-JAN-2003; 2003WO-US0031194.  
XX 31-JAN-2002; 2002US-000600087.  
XX 15-MAR-2002; 2002US-0364045P.  
XX 15-MAR-2002; 2002US-0364055P.  
XX 30-DEC-2002; 2002US-0436643P.  
XX (GENE-) GENE LOGIC INC.  
XX Mendrick D, Porter M, Johnson K, Higgs B, Castle A, Elashoff M;  
XX WPI; 2003-689530/65.  
XX Predicting a toxic effect of a compound, useful in identifying toxicity  
XX markers in liver tissues or cells for drug screening and toxicity assays,  
XX comprises preparing gene expression profile of tissue or cells exposed to  
XX the compound.  
XX Claim 1; SEQ ID NO 2776; 1156pp; English.  
XX The present invention relates to a method for predicting a toxic effect  
XX of a compound. The method comprises preparing a gene expression profile  
XX of a tissue or cell sample exposed to the compound, and comparing the  
XX gene expression profile to a database comprising SEQ ID 1-4925, where  
XX differential expression of the gene indicates at least one toxic effect.  
XX The method is useful for predicting at least one toxic effect of a  
XX compound, predicting hepatotoxicity or the progression of a toxic effect  
XX of a compound, identifying an agent that modulates the onset or  
XX progression of a toxic response, predicting the cellular pathways that a  
XX compound modulates in a cell, and identifying an agent that modulates at  
XX least one activity of a protein. The method and compositions of the  
XX present invention using a database of genes having liver toxin-induced  
XX differential expression, are useful in identifying toxicity markers in  
XX liver tissues or cells for drug screening and toxicity assays. Note: The  
XX sequence data for this patent did not form part of the printed  
XX specification, but was obtained in electronic format directly from WIPO  
XX at ftp.wipo.int/pub/published\_pct\_sequences.  
SQ Sequence 540 BP; 162 A; 119 C; 85 G; 174 T; 0 U; 0 Other;

Query Match 3.3%; Score 181.6; DB 10; Length 540;  
Best Local Similarity 75.8%; Pred. No. 6.8e-36;  
Matches 388; Conservative 0; Mismatches 69; Indels 55; Gaps 11;  
QY 4960 ATATTATAAAATTTTCAGATTACCTTCACTATCAAAATGAGTAATGAC----- 5011  
DB 499 ATGTTGAAATTTGTTCCAGATTATACCTTCACTATCAAAATGAGTAATGAC 440  
QY 5012 ----TCATGCTGCAGACATGCTCTAGTGGTGGCAGACAGAGGATCT-----TTG 5060

```
Db 439 ATAATCAGCGCTCAGAGATGTCCTCAATGGTGGCAAGACAGAGAGGCTCAAACTGAAGG 380
Qy 5061 ACTAAGGAGAAAAAATGTCATTCATCCAGCCGCCAGGAAAGAACACCTCCAAAGGCA 5120
Db 379 AGAAAAACAATTAATCTGTCATTTGTCAT-CGATCCCAAGGAAGAACACCTCCGAGGCA 321
Qy 5121 GGCAGGAGGAGCAGGAGCATGTTGTTCTAGTTGAATACACATTC-AGTCTTTGCAAGTG 5179
Db 320 -----AGGCAGGCATGGCGGTTCTAGCTGAACACTTACACTCAGTCTTGCAGTG 272
Qy 5180 GTGCTTTT--AGATCTGTGTAGCATGTGAGGCTCTGTACAGTGGGG--GCCACACTTCTGA 5236
Db 271 GTGCTTTTAGAGATCTCTGTAGCATGTGAGGCTCTGTACAGTGGGGACCCACACTTCCGA 212
Qy 5237 GGGCTGAAATGTGCAACCCCTTTATCTAACTTGAATCAAAACCGTCAAAATTTTATTTTT 5296
Db 211 GTGCTGAAATGTGCAACCC-THATCTAACTTGAATCAAAACCGTCAAGATTTTATTTTT 153
Qy 5297 TATAATTTAAGAA-----AGAGTTGGGAATGACATTTTTTTGAGTTGGCCCTTTTCAGCTCA 5352
Db 152 TATAATTTAAGAAAGTAGAGACGGAATGAGAAATTTTTTTTTTTTGGAAATGAGATGTT 93
Qy 5353 G-----TCATTTACGTGAAC--GTGGAGATTTGATAGCTCAGATTATTT 5397
Db 92 GAGTTGGTTTCAGATCAATTTTACATGTAACTAGTGTAGATTTTATAGCGGAGATTATTT 33
Qy 5398 TGTATATAATTTATTAACCTAATCTGTAATTTGT 5429
Db 32 TGTATATAATTTATTAACCTAATCTGTAATTTGT 1
```

```
RESULT 9
ABX52167
ID ABX52167 standard; cDNA; 399 BP.
XX
AC ABX52167;
XX
DT 25-FEB-2003 (first entry)
XX
DE Bovine EST associated with lactation/muscle/fat deposition #2096.
XX
KW Bovine; ss; EST; expressed sequence tag; lactation; LMFD;
KW muscle deposition; fat deposition; genome mapping; gene identification;
KW gene analysis; cattle breeding.
XX
OS Bos Taurus.
XX
PN US2002137160-A1.
XX
PD 26-SEP-2002.
XX
PF 26-OCT-2001; 2001US-00983965.
XX
PR 17-DEC-1998; 98US-0113678P.
XX
PR 15-DEC-1999; 99US-00465231.
XX
PA (BYAT/) BYATT J C.
PA (MATH/) MATHIALAGAN N.
PA (TAON/) TAO N.
PA (WARR/) WARREN W C.
XX
PI Byatt JC, Mathialagan N, Tao N, Warren WC;
XX
PI WPT; 2003-102386/09.
XX
PT Purified nucleic acid molecules, useful for genome mapping, gene
PT identification and analysis, cattle breeding or preparation of constructs
PT for cattle gene expression and genetically improved cattle.
XX
PS Claim 2; SEQ ID NO 2096; 38pp; English.
XX
CC The invention relates to a purified nucleic acid molecule associated with
```

```
CC lactation or muscle and fat deposition (designated LMFD), derived from
CC cattle, and the LMFD nucleic acid can specifically hybridize to a second
CC nucleic acid molecule comprising any of 5912 nucleotide sequences,
CC appearing as ABX50072-ABX5983, or complements of them. Also included are
CC : (1) a transformed cell having a nucleic acid comprising an LMFD nucleic
CC acid linked to a promoter and a 3' non-translated sequence that
CC functions in the cell to cause termination of transcription and addition
CC of polyadenylated ribonucleotides to a 3' end of the mRNA molecule; and
CC (2) determining a level or pattern of a molecule in a bovine cell or
CC tissue comprising: (a) incubating a marker nucleic acid (comprising any
CC of the 5912 nucleic acid sequences or its complement or fragment) with a
CC complementary nucleic acid molecule obtained from the bovine cell or
CC tissue, where hybridisation between the marker nucleic acid and the
CC complementary nucleic acid permits the detection of the molecule; and (b)
CC detecting the level or pattern of the complementary nucleic acid, where
CC the detection of the complementary nucleic acid is predictive of the
CC level or pattern of the molecule. The LMFD nucleic acid is used for
CC determining a level or pattern of a molecule in a bovine cell or tissue.
CC It is useful for genome mapping, gene identification and analysis, cattle
CC breeding, preparation of constructs for use in cattle gene expression, or
CC for genetically improving cattle. The present sequence is one of the 5912
CC bovine LMFD EST (expressed sequence tag) nucleic acids. Note: The present
CC sequence was not shown in the specification but was obtained in
CC electronic format from the USPTO web site:
CC seqdata.uspto.gov/sequence.html?DocID=20020137160
XX
SQ Sequence 399 BP; 133 A; 62 C; 82 G; 122 T; 0 U; 0 Other;
```

```
Query Match 3.1%; Score 170.8; DB 8; Length 399;
Best Local Similarity 79.1%; Pred. No. 3.5e-33;
Matches 231; Conservative 0; Mismatches 52; Indels 9; Gaps 2;
```

```
Qy 2330 AAAGGTAACCTTCCTCGAAGAAGTTACTGAGTATTATATAAGTGGTGATGAGATCGCAA 2389
Db 102 ATAGGTAACCTTCCTCGAAGAAGTTACTGAGTATTATATAAGCGGTGATGAGATCGAAA 161
Qy 2390 AGGACCATGGGAAGAAATTTGCAAGGATGATGATGATGATGATGATGATGATGATGATG 2449
Db 162 AGGACCATGGGAAGAAATTTGCAAGGATGATGATGATGATGATGATGATGATGATGATG 221
Qy 2450 AGAAGTTGCCATTTGGCTACTGCTTGGCTTTGAGCACAGAGAAAAA-----ATGTTTA 2502
Db 222 AGAAGAAAGCTAATTTGGATCTGCTTTGACATTTGAGCACAGAGAAAAAAGATGTTTAATA 281
Qy 2503 ATAGACTGAGGATCGAGTCAAGGACTTACTGTTGTACAGCAATGTTAAGA--AGTGAAC 2560
Db 282 CTTCAAGCAACATGTTTCAAGGACTTAATGTTTTCGAGCAATGTTAAGATGATTCGAC 341
Qy 2561 AGCCTGCAACCCGTGCCACTCTGCTCTTACTTTGAGAGTTTCCCTTAAAAA 2612
Db 342 AGCCTCTGGCCCTAGCATACACTACCTCTTACTTTGAGAGGGTACTTTTAAAA 393
```

```
RESULT 10
AAI96694/c
ID AAI96694 standard; cDNA; 689 BP.
```

```
XX
AC AAI96694;
```

```
XX
DT 13-NOV-2001 (first entry)
```

```
XX
DE Human neuroblastoma expressed polynucleotide SEQ ID NO 2769.
```

```
XX
KW Human; neuroblastoma; malignancy; cancer; tumour marker; N-myc; TrkA; ss.
```

```
XX
OS Homo sapiens.
```

```
XX
PN WO200166719-A1.
```

```
XX
PD 13-SEP-2001.
```

```
XX
PF 02-MAR-2001; 2001WO-JP001629.
```

```
XX
```



XX OS Homo sapiens.  
 XX PN EP1074617-A2.  
 XX XX  
 XX PD 07-FEB-2001.  
 XX XX  
 XX PF 28-JUL-2000; 2000EP-00116126.  
 XX XX  
 XX PR 29-JUL-1999; 99JP-00248036.  
 XX PR 27-AUG-1999; 99JP-00300253.  
 XX PR 11-JAN-2000; 2000JP-00118776.  
 XX PR 02-MAY-2000; 2000JP-00183767.  
 XX PR 09-JUN-2000; 2000JP-00241899.  
 XX XX  
 XX PA (HELI-) HELIX RES INST.  
 XX XX  
 XX PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
 XX PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
 XX XX  
 XX DR WPI; 2001-318749/34.  
 XX XX  
 XX PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.  
 XX PT  
 XX PS Claim 3; SEQ ID NO 8501; 2537pp + Sequence Listing; English.  
 XX XX  
 XX CC The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention  
 XX CC  
 XX SQ Sequence 543 BP; 189 A; 104 C; 97 G; 149 T; 0 U; 4 Other;  
 XX XX  
 XX Query Match 2.0%; Score 109; DB 4; Length 543;  
 XX Best Local Similarity 63.7%; Pred. No. 3.8e-17;  
 XX Matches 254; Conservative 0; Mismatches 119; Indels 26; Gaps 5;  
 XX XX  
 XX QY 2407 TTGCAGGGATGGATCCAGGTTCCAGAACCAATTCAGAACAGAGATGGCCATTCGCT 2466  
 XX DB 543 TTTTCAAGGATGAATCCAGTTCCNAGAAAGGATTCACAGAACAGAGATGCTATTGGAT 484  
 XX QY 2467 ACTGCTTGGCCCTTTGAGCACAGAGAAAATGTTTAATAGACT----GAGGATCGAGTCA 2522  
 XX DB 483 ATGGCTTGACATTTGACCCAGAGAAAGATGTTTATAGACTCCAGGAACATGCTTCA 424  
 XX QY 2523 AAGGACTTACTGTTGTACAGCAATGTTTAAGAAAGTGA-ACAGCTTGCAACCCGTCGCCACT 2591  
 XX DB 423 AAGGACTTAATGTTCTCAAGCAATGTTGAGTTGGCAGCCTGTAGTCTCTAGCTAGCATACA 364  
 XX QY 2592 CTGTCTCTTACTTGAGA-GTTTCCCTTTAAACAAACACATGGCAGCTGCTCTTGACATG 2640  
 XX DB 363 CTACCTCTTACCTGAGAGGTGCTCTTTTAAACAAATCTTGGCAGCTGCTCTTGACATT 304

QY 2641 TTTTAAAGAAACAACCTTGATCTAGAGATGCGAGTTTGATATTATTTGGTAATGTGCT 2700  
 DB 303 TTTTATTTAGAGGAAATGTAACCTGGATCTAGTTTAAATTTTATTTTTCACATATCC 244  
 QY 2701 CATTAGAA-----ACACCAACTCCGATAAAGCAAGAT-----CTCTTATC 2740  
 DB 243 CACTCAGAACATTCAGGTTTGAAGCCGCGCTGATAAAGGATGAAGTACTAGTGATT 184  
 QY 2741 TGTAAATCCTCTCTTTTCTTATTTAGTTAGTGGTGGTTT 2779  
 DB 183 TCTAATCCTCCCTTTTGTGTTTATTTAGTTAGTGGTGGTTT 145  
 RESULT 13  
 AAF21761  
 ID AAF21761 standard; DNA; 1028 BP.  
 XX  
 AC AAF21761;  
 XX  
 DT 27-MAR-2001 (first entry)  
 XX  
 DE Human breast and ovarian cancer associated antigen gene SEQ ID 148.  
 XX  
 KW Human; breast cancer; ovarian cancer; cytostatic; immunosuppressive; neurotropic; neuroprotective; antiviral; antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor; antitumor; anticonvulsant; antibacterial; antifungal; antiparasitic; cardiac; immune disorder; Addison's disease; allergy; autoimmune haemolytic anaemia; autoimmune thyroiditis; diabetes mellitus; Crohn's disease; multiple sclerosis; rheumatoid arthritis; ulcerative colitis; cardiovascular disorder; wound healing; neurological disease; ds.  
 KW  
 XX Homo sapiens.  
 XX  
 PN WO200055173-A1.  
 XX  
 PD 21-SEP-2000.  
 XX  
 PF 08-MAR-2000; 2000WO-US005881.  
 XX  
 PR 12-MAR-1999; 99US-0124270P.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Rosen CA, Ruben SM;  
 XX  
 DR WPI; 2000-611515/58.  
 DR P-PSDB; AAB58858.  
 XX  
 PT New human breast and ovarian cancer associated gene sequences and the polypeptides encoded by these genes, useful in the prevention, treatment and diagnosis of cancer, immune disorders, cardiovascular disorders and neurological diseases.  
 PT  
 PT Claim 1; Page 592; 1299pp; English.  
 XX  
 PS  
 XX  
 XX Sequences AAF21614 - AAF22031 represent DNA sequences encoding human proteins AAB58711 - AAB59128. The DNA and protein sequences are associated with breast and ovarian cancer. Included in the invention are sequences AAF22032 - AAF22040 and AAB59129 which are used in the isolation and characterisation of the DNA and protein sequences of the invention. The breast and ovarian cancer associated DNA, protein, agonist or antagonist sequences exhibit cytostatic; immunosuppressive; neurotropic; neuroprotective; antiviral; antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor; antitumor; anticonvulsant; antibacterial; antifungal; antiparasitic and cardiac activity. The polynucleotide and protein sequences are used in the diagnosis of cancer, particularly breast and ovarian cancer. The nucleic acid sequences, proteins, agonists and agonists may also be used in the diagnosis, prevention and treatment of immune disorders e.g. Addison's disease, allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative colitis;  
 CC

CC cardiovascular disorders such as myocardial ischaemias; wound healing;  
 CC neurological diseases such as cerebral anoxia and epilepsy; and  
 CC infectious diseases  
 XX  
 SQ Sequence 1028 BP; 320 A; 174 C; 199 G; 335 T; 0 U; 0 Other;  
 Query Match 1.8%; Score 99.6; DB 3; Length 1028;  
 Best Local Similarity 70.8%; Pred. No. 1.5e-14;  
 Matches 177; Conservative 0; Mismatches 64; Indels 9; Gaps 3;  
 QY 5214 ACAGTGGGGCCACACTCTGAGGGCTGAATGTGGCAACCCCTTATC--TAACCTGAA 5271  
 DB 757 ACAGTAAAGTTCACATCTCTGAGTGATGAATGTAAACACTTCTTCATCTTTAACTTGA 816  
 QY 5272 ATCAAAACCGTCAAAATTTTATTTT-----TATAATTTAAGAAAGAGTTCGGGAATGAC 5325  
 DB 817 ATCAAACTATCAGATTTTATTTTGTATTAATTTAAGGAAGTAAAGTTAGGGGACTAGA 876  
 QY 5326 ATTTTTCAGTTGGCTTTTTCAGCTCAGTCATTT--TACGTGTAACGTGGAGATTGTAG 5384  
 DB 877 AGACTCTAAATTGGCTTCTACAGATCAATAATTTAAATGTAAGTGTGGGATTTTATAG 936  
 QY 5385 CTCAGATTATTTGTATATAATTAATTAATTAATCTGTAATTTGTAATAATATATTGTC 5444  
 DB 937 TTAATAATTTATTTGTATATAACATAAATACTGTAATTTGTAATAATATATTGTC 996  
 QY 5445 AATTATTAAA 5454  
 DB 997 AATTATTAAA 1006  
 RESULT 14  
 AAH13104/c  
 ID AAH13104 standard; cDNA; 482 BP.  
 XX  
 AC AAH13104;  
 XX  
 DT 26-JUN-2001 (first entry)  
 XX  
 DE Human cDNA clone (3'-primer) SEQ ID NO:9939.  
 XX  
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN EP1074617-A2.  
 XX  
 PD 07-FEB-2001.  
 XX  
 PF 28-JUL-2000; 2000EP-00116126.  
 XX  
 PR 29-JUL-1999; 99JP-00248036.  
 PR 27-AUG-1999; 99JP-00300253.  
 PR 11-JAN-2000; 2000JP-00118776.  
 PR 02-MAY-2000; 2000JP-00183767.  
 PR 09-JUN-2000; 2000JP-00241899.  
 XX  
 PA (HELI-) HELIX RES INST.  
 XX  
 XX Ota T, Isegai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
 XX  
 DR WPI; 2001-318749/34.  
 XX  
 PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.  
 XX  
 PS Claim 3; SEQ ID NO 9939; 2537pp + Sequence Listing; English.  
 XX  
 XX The present invention describes primer sets for synthesising 5602 full-length cDNAs defined in the specification. Where a primer set comprises:

CC (a) an oligo-dT primer and an oligonucleotide complementary to the  
 CC complementary strand of a polynucleotide which comprises one of the 5602  
 CC nucleotide sequences defined in the specification, where the  
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
 CC of an oligonucleotide comprising a sequence complementary to the  
 CC complementary strand of a polynucleotide which comprises a 5'-end  
 CC sequence and an oligonucleotide comprising a sequence complementary to a  
 CC polynucleotide which comprises a 3'-end sequence, where the  
 CC oligonucleotide comprises at least 15 nucleotides and the combination of  
 CC the 5'-end sequence/3'-end sequence is selected from those defined in the  
 CC specification. The primer sets can be used in antisense therapy and in  
 CC gene therapy. The primers are useful for synthesising polynucleotides,  
 CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAH95893  
 CC represent human amino acid sequences; and AAH13629 to AAH13632 represent  
 CC oligonucleotides, all of which are used in the exemplification of the  
 CC present invention  
 XX  
 SQ Sequence 482 BP; 170 A; 83 C; 52 G; 170 T; 0 U; 7 Other;  
 Query Match 1.8%; Score 98; DB 4; Length 482;  
 Best Local Similarity 70.4%; Pred. No. 2.4e-14;  
 Matches 176; Conservative 0; Mismatches 65; Indels 9; Gaps 3;  
 QY 5214 ACAGTGGGGCCACACTCTGAGGGCTGAAATGTGGCAACCCCTTATC--TAACCTGAA 5271  
 DB 261 ACAGTAAAGTTCACATCTCTGAGTGATGAATGTAAACACTTCTTCATCTTTAACTTGA 202  
 QY 5272 ATCAAAACCGTCAAAATTTTATTTT-----TATAATTTAAGAAAGTTCGGGAATGAC 5325  
 DB 201 ATCAAACTATCAGATTTTATTTTGTATTAATTTAAGGAAGTAAAGTTAGGGGACTAGA 142  
 QY 5326 ATTTTTCAGTTGGCTTTTTCAGCTCAGTCATTT--TACGTGTAACGTGGAGATTGTAG 5384  
 DB 141 AGACTCTAAATTTGGCTTCTACAGATCAATAATTTAAATGTAAGTGTGGGATTTTATAG 82  
 QY 5385 CTCAGATTATTTGTATATAATTAATTAATTAATCTGTAATTTGTAATAATATATTGTC 5444  
 DB 81 TTAATAATTTATTTGTGTATATAACATAATTAATCTGTAATTTGTAATAATATATTGTC 22  
 QY 5445 AATTATTAAA 5454  
 DB 21 AATTATTAAA 12  
 RESULT 15  
 AAH16576  
 ID AAH16576 standard; cDNA; 2148 BP.  
 XX  
 AC AAH16576;  
 XX  
 DT 26-JUN-2001 (first entry)  
 XX  
 DE Human cDNA sequence SEQ ID NO:15655.  
 XX  
 XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.  
 KW Homo sapiens.  
 OS  
 XX  
 PN EP1074617-A2.  
 XX  
 PD 07-FEB-2001.  
 XX  
 PF 28-JUL-2000; 2000EP-00116126.  
 XX  
 PR 29-JUL-1999; 99JP-00248036.  
 PR 27-AUG-1999; 99JP-00300253.  
 PR 11-JAN-2000; 2000JP-00118776.  
 PR 02-MAY-2000; 2000JP-00183767.  
 PR 09-JUN-2000; 2000JP-00241899.  
 XX





**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 15, 2005, 12:27:30 ; Search time 815.974 Seconds  
(without alignments)  
10965.010 Million cell updates/sec

Title: US-10-650-482-3

Perfect score: 5468

Sequence: 1 cggctctccgtctgcgcctg.....attaaaaaaaaaaaaa 5468

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : Issued Patents\_NA.\*

1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/1/ina/PTCUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	93.8	1.7	218	4	US-09-513-999C-33878
2	78.8	1.4	3046	1	US-08-726-725-1
3	75.2	1.4	2070	4	US-10-329-668-5
4	74.8	1.4	2010	4	US-09-809-545A-52
5	74.2	1.4	13011	2	US-08-791-849A-14
6	73.6	1.3	7970	1	US-08-135-511-31
7	73.6	1.3	7997	1	US-08-187-453-31
8	73.6	1.3	7997	1	US-08-562-985A-7
9	71.2	1.3	16011	4	US-09-600-319-3
10	69.8	1.3	16011	4	US-09-600-319-3
11	69.6	1.3	2432	3	US-08-974-022-1
12	69.6	1.3	2432	3	US-08-795-445A-1
13	69.6	1.3	2432	3	US-08-795-447A-1
14	69.6	1.3	2432	3	US-08-974-186-1
15	69.6	1.3	2432	3	US-08-795-446B-1
16	69.6	1.3	2432	3	US-08-706-945D-123
17	69.6	1.3	2432	3	US-08-577-788C-1
18	69.4	1.3	5109	3	US-08-930-055A-2
19	68	1.2	2226	2	US-08-031-538-1
20	68	1.2	3963	1	US-08-464-961-1
21	68	1.2	3963	3	US-08-907-800A-1
22	68	1.2	3963	3	US-08-983-541-1
23	68	1.2	3963	3	US-08-969-315-1
24	68	1.2	3963	5	PTC-US96-08233-1
25	67.6	1.2	4634	4	US-09-244-805-58
26	67.4	1.2	2614	4	US-09-052-753B-8
27	67.4	1.2	5798	4	US-09-377-285B-33
28	66.2	1.2	1970	4	US-09-621-781B-1
29	66	1.2	5305	4	US-09-244-805-34
30	64.2	1.2	1900	3	US-09-604-978-3
31	64.2	1.2	1900	4	US-09-604-728-3
32	64.2	1.2	1900	4	US-10-325-878-3
33	62.4	1.1	2121	3	US-09-117-250-2
34	59.4	1.1	1046	4	US-09-509-712B-113
35	54.8	1.0	1824	4	US-09-244-805-5
36	54.6	1.0	7218	1	US-08-232-463-14
37	54.4	1.0	9792	4	US-09-635-872A-14
38	54.4	1.0	9792	4	US-09-636-077A-14
39	54.4	1.0	9792	4	US-09-636-060C-14
40	54.4	1.0	9792	4	US-09-586-552-14
41	54.4	1.0	9792	4	US-09-636-596C-14
42	54.4	1.0	9792	4	US-10-023-894-21
43	54.4	1.0	9792	4	US-10-306-686-14
44	53.6	1.0	3046	1	US-08-726-725-1
45	52.4	1.0	7218	1	US-08-232-463-14
46	51.6	0.9	2070	4	US-10-329-668-5
47	50.2	0.9	2596	4	US-09-554-726A-8
48	49.8	0.9	5935	3	US-09-178-973B-17
49	49.8	0.9	5935	3	US-09-419-568F-29
50	49.8	0.9	5935	3	US-09-354-243B-29
51	49.6	0.9	5687	2	US-08-380-403A-3
52	49.6	0.9	5687	2	US-08-895-628-3
53	49.6	0.9	5687	3	US-08-895-810D-3
54	49.4	0.9	7970	1	US-08-135-511-31
55	49.4	0.9	7997	1	US-08-187-453-31
56	49.4	0.9	7997	1	US-08-562-985A-7
57	49	0.9	5798	4	US-09-377-285B-33
58	48.8	0.9	30310	4	US-09-657-346A-96
59	48.8	0.9	30310	4	US-09-657-346A-96
60	48.6	0.9	14507	3	US-08-785-150-1
61	48.6	0.9	14507	3	US-09-660-299-1
62	48.6	0.9	14507	3	US-09-435-377-1
63	48.6	0.9	14507	4	US-09-373-928-1
64	48.2	0.9	2010	4	US-09-809-545A-52
65	47.8	0.9	612	4	US-09-902-540-1357
66	47.6	0.9	5191	1	US-08-340-428B-1
67	47.6	0.9	5191	5	PCT-US93-07306-1
68	47.2	0.9	2044	2	US-09-327-138C-38
69	47.2	0.9	2226	2	US-08-031-538-1
70	47.2	0.9	2863	4	US-09-327-138C-2
71	47.2	0.9	3646	4	US-09-327-138C-36
72	47.2	0.9	3665	4	US-09-327-138C-35
73	47.2	0.9	3686	4	US-09-327-138C-3
74	47.2	0.9	3686	4	US-09-327-138C-37
75	47.2	0.9	3940	4	US-09-327-138C-1
76	46.8	0.9	1039	4	US-09-902-540-1280
77	46.8	0.9	9840	3	US-09-534-638-1
78	46.8	0.9	48974	3	US-08-920-422-17
79	46.4	0.8	817	4	US-09-809-545A-7
80	46.4	0.8	4403765	3	US-09-103-840A-2
81	46.2	0.8	6727	3	US-08-629-643A-5
82	46.2	0.8	6727	3	US-09-280-799-1
83	46.2	0.8	6727	3	US-09-155-884-5
84	46.2	0.8	51259	3	US-08-781-891-209
85	46.2	0.8	51259	4	US-09-618-166-209
86	46.2	0.8	4403765	3	US-09-103-840A-2
87	46.2	0.8	4411529	3	US-09-103-840A-1
88	46	0.8	12141	3	US-09-488-671-10
89	45.8	0.8	2432	3	US-08-974-022-1
90	45.8	0.8	2432	3	US-08-795-445A-1
91	45.8	0.8	2432	3	US-08-795-447A-1
92	45.8	0.8	2432	3	US-08-974-186-1
93	45.8	0.8	2432	3	US-08-795-446B-1
94	45.8	0.8	2432	3	US-08-706-945D-123
95	45.8	0.8	2432	4	US-08-577-788C-1
96	45.8	0.8	37950	3	US-09-338-907-183
97	45.8	0.8	37950	3	US-09-218-207-183
98	45.4	0.8	601	4	US-09-949-016-60008
99	45.4	0.8	601	4	US-09-949-016-60009
100	45.4	0.8	343352	4	US-09-949-016-13498

Sequence 1, Appli  
Sequence 34, Appli  
Sequence 3, Appli  
Sequence 3, Appli  
Sequence 2, Appli  
Sequence 113, App  
Sequence 5, Appli  
Sequence 14, Appli  
Sequence 14, Appli  
Sequence 14, Appli  
Sequence 14, Appli  
Sequence 14, Appli  
Sequence 14, Appli  
Sequence 21, Appli  
Sequence 14, Appli  
Sequence 14, Appli  
Sequence 14, Appli  
Sequence 5, Appli  
Sequence 8, Appli  
Sequence 17, Appli  
Sequence 29, Appli  
Sequence 29, Appli  
Sequence 3, Appli  
Sequence 3, Appli  
Sequence 31, Appli  
Sequence 7, Appli  
Sequence 32, Appli  
Sequence 96, Appli  
Sequence 96, Appli  
Sequence 1, Appli  
Sequence 1, Appli  
Sequence 1, Appli  
Sequence 1, Appli  
Sequence 52, Appli  
Sequence 1357, Ap  
Sequence 1, Appli  
Sequence 1, Appli  
Sequence 1, Appli  
Sequence 38, Appli  
Sequence 1, Appli  
Sequence 2, Appli  
Sequence 1280, Ap  
Sequence 1, Appli  
Sequence 17, Appli  
Sequence 2, Appli  
Sequence 5, Appli  
Sequence 5, Appli  
Sequence 209, App  
Sequence 209, App  
Sequence 2, Appli  
Sequence 1, Appli  
Sequence 10, Appli  
Sequence 1, Appli  
Sequence 1, Appli  
Sequence 1, Appli  
Sequence 1, Appli  
Sequence 123, App  
Sequence 183, App  
Sequence 183, App  
Sequence 60008, A  
Sequence 60009, A  
Sequence 13498, A

ALIGNMENTS

RESULT 1  
US-09-513-999C-33878  
; Sequence 33878, Application US/09513999C  
; Patent No. 6783961  
; GENERAL INFORMATION:  
; APPLICANT: Dumas Milne Edwards, J.B.  
; APPLICANT: Duclert, A.  
; APPLICANT: Giordano, J.Y.  
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.  
; FILE REFERENCE: 59.US2.REG  
; CURRENT APPLICATION NUMBER: US/09/513,999C  
; CURRENT FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/122,487  
; PRIOR FILING DATE: 1999-02-26  
; NUMBER OF SEQ ID NOS: 36681  
; SOFTWARE: Patent.pm  
; SEQ ID NO 33878  
; LENGTH: 218  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-513-999C-33878

Query Match 1.7%; Score 93.8; DB 4; Length 218;  
Best Local Similarity 70.6%; Pred. No. 1.8e-15;  
Matches 125; Conservative 0; Mismatches 52; Indels 0; Gaps 0;  
QY 1550 ACAGAGCCCCCAGGAGCAGTCTGTTTGTGAATTACCGTGGAAGAAATGTGAAGA 1609  
Db 42 AGAGAGCCCTTCTGAGGGCTGTCCATCTAGTGAGATACCTATGGAAGAGCCTGGAGA 101  
QY 1610 GGACCACACTAATGCAACTGACCTCTCAGATAGAGGAGAGAGCCCTTCTGTTTCTACCA 1669  
Db 102 GGGCCGAATAGTGTAGTGTGTTACTCATACCTAGAGGTGACCTTCCCATTTCTGCCAG 161  
QY 1670 ACCAGTTGTAGCAACAACTGATAGATATATTTTGGAGCGGCCCCAGTACT 1726  
Db 162 ACCAGCTTGTAGTACAACTGATAGATATATTTTGGAGGTGCATCCAGTACCT 218

RESULT 2  
US-08-726-725-1/c  
; Sequence 1, Application US/08726725  
; Patent No. 5773290  
; GENERAL INFORMATION:  
; APPLICANT: Gould, Michael N.  
; APPLICANT: Chen, Kai-Shun  
; TITLE OF INVENTION: MAMMARY GLAND-SPECIFIC PROMOTERS  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Quarles & Brady  
; STREET: 411 East Wisconsin Avenue  
; CITY: Milwaukee  
; STATE: Wisconsin  
; COUNTRY: U.S.A.  
; ZIP: 53202-4497  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/726,725  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Baker, Jean C.  
; REGISTRATION NUMBER: 35,433

; REFERENCE/DOCKET NUMBER: 960296.93863  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (414) 277-5709  
; TELEFAX: (414) 271-3552  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3046 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-08-726-725-1

Query Match 1.4%; Score 78.8; DB 1; Length 3046;  
Best Local Similarity 71.2%; Pred. No. 1.6e-10;  
Matches 104; Conservative 0; Mismatches 42; Indels 0; Gaps 0;  
QY 3613 TTTTCTCCCTCTTTTTTTTTTCCCCCAGAGCTAGGACCAAAACCCAGGCTTTGCACTT 3672  
Db 421 TTTTTTTTTTTTTTTTTTTTTTTCGGAGCTGGGACCGAACCCAGGGCTTGGCGCTT 362  
QY 3673 GCTAGGCAAGCGCTCTACCACTAGCTAAATCCCAAAACCAACCCCGCTTTTCTTTT 3732  
Db 361 CCTAGGTAAGCGCTCTGCCACTGAGCTAAATCCCAACCCCGCTACTTCTTTTATTAT 302  
QY 3733 TTTAAAGACATGGTCTTTATATAGTCT 3758  
Db 301 TTATTATTATTGTATGAGGTCT 276

RESULT 3  
US-10-329-668-5  
; Sequence 5, Application US/10329668  
; Patent No. 6696473  
; GENERAL INFORMATION:  
; APPLICANT: Martin Richard  
; APPLICANT: Brenton Todd  
; APPLICANT: Kahl Jeffrey Dean  
; APPLICANT: Wang Tie-Lin  
; TITLE OF INVENTION: HETEROCYCLIC MODULATORS OF NUCLEAR RECEPTORS  
; FILE REFERENCE: 38205-3001  
; CURRENT APPLICATION NUMBER: US/10/329,668  
; CURRENT FILING DATE: 2002-12-20  
; PRIOR APPLICATION NUMBER: 60/342,720  
; PRIOR FILING DATE: 2001-12-21  
; NUMBER OF SEQ ID NOS: 18  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 5  
; LENGTH: 2070  
; TYPE: DNA  
; ORGANISM: Rattus norvegicus  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (172)...(1581)  
; PUBLICATION INFORMATION:  
; DATABASE ACCESSION NUMBER: GeneBank U18374  
; DATABASE ENTRY DATE: 1995-06-21  
US-10-329-668-5

Query Match 1.4%; Score 75.2; DB 4; Length 2070;  
Best Local Similarity 76.7%; Pred. No. 1.2e-09;  
Matches 92; Conservative 0; Mismatches 28; Indels 0; Gaps 0;  
QY 3593 TGTGTCACTTCAGTGGTGGATTTTCTCCCTCTTTTTTTTCCCCCAGAGCTGAGGACC 3652  
Db 1896 TTATCAAAATAGCTGTTTCTCTCTTTTTTTTTTTTTTTTTCGGAGCTGGGACT 1955  
QY 3653 AAACCCAGGGCTTTGCACTTTGCTAGGCAAGCGCTCTACCACTGAGCTAAATCCCAAC 3712  
Db 1956 GAACCCAGGGCTTGGCTTGTCTAGGCAAGCGCTCTACCACTGAGCTAAATCCCAAC 2015

RESULT 4







	D	b	4888	G G G C T T G C G C T T G C T A G G C A G T G C T C T A C C A C T G A G C T A A A T C C C C A G C C C C A G C A C T	4829
	Q	y	3720	C C C C T T T T 3727 	
	D	b	4828	A A A G T T T T 4821	
RESULT 9					
US-09-600-319-3/c					
; Sequence 3, Application US/09600319					
; Patent No. 6780610					
; GENERAL INFORMATION:					
; APPLICANT: Owens, Gary					
; APPLICANT: Madsen, Cort					
; TITLE OF INVENTION: Identification of a Smooth Muscle Cell (SMC) Specific Smooth Musc					
; TITLE OF INVENTION: Myosin Heavy Chain (SM-MHC) Promoter/Enhancer					
; FILE REFERENCE: 00241-03					
; CURRENT APPLICATION NUMBER: US/09/600,319					
; CURRENT FILING DATE: 2001-10-11					
; PRIOR APPLICATION NUMBER: PCT/US99/01038					
; PRIOR FILING DATE: 1999-01-15					
; PRIOR APPLICATION NUMBER: 60/071,300					
; PRIOR FILING DATE: 1998-01-16					
; NUMBER OF SEQ ID NOS: 3					
; SOFTWARE: PatentIn version 3.0					
; SEQ ID NO 3					
; LENGTH: 16011					
; TYPE: DNA					
; ORGANISM: Rattus norvegicus					
US-09-600-319-3					
Query Match 1.3%; Score 71.2; DB 4; Length 16011;					
Best Local Similarity 75.9%; Pred. No. 6.4e-08;					
Matches 88; Conservative 0; Mismatches 28; Indels 0; Gaps 0;					
	Q	y	3612	A T T T T C C C C T C T T T T T T T T T T T T T T T T C C C C C A G A G C T G A G G A C C A A A C C C A G G C T T T G C A C T	3671
	D	b	14537	A T C G A G C T G G G G A C C G A A C C C A G G C C T T G C G C T	14478
	Q	y	3672	T G C T A G G A A G C G C T C T A C C A C A C T G A G C T A A A T C C C A A A C C A C C C C C C C C C T T T T	3727
	D	b	14477	T C T A G T A A G C G C T C T A C C A C T G A G C T A A A T C C C A G C C C C A G A C C C T T C A T T T	14422
RESULT 10					
US-09-600-319-3					
; Sequence 3, Application US/09600319					
; Patent No. 6780610					
; GENERAL INFORMATION:					
; APPLICANT: Owens, Gary					
; APPLICANT: Madsen, Cort					
; TITLE OF INVENTION: Identification of a Smooth Muscle Cell (SMC) Specific Smooth Musc					
; TITLE OF INVENTION: Myosin Heavy Chain (SM-MHC) Promoter/Enhancer					
; FILE REFERENCE: 00241-03					
; CURRENT APPLICATION NUMBER: US/09/600,319					
; CURRENT FILING DATE: 2001-10-11					
; PRIOR APPLICATION NUMBER: PCT/US99/01038					
; PRIOR FILING DATE: 1999-01-15					
; PRIOR APPLICATION NUMBER: 60/071,300					
; PRIOR FILING DATE: 1998-01-16					
; NUMBER OF SEQ ID NOS: 3					
; SOFTWARE: PatentIn version 3.0					
; SEQ ID NO 3					
; LENGTH: 16011					
; TYPE: DNA					
; ORGANISM: Rattus norvegicus					
US-09-600-319-3					
Query Match 1.3%; Score 69.8; DB 4; Length 16011;					
Best Local Similarity 86.5%; Pred. No. 1.6e-07;					
Matches 77; Conservative 0; Mismatches 12; Indels 0; Gaps 0;					
	Q	y	3624	T T T T T T T T T T T C C C C C A G A G C T G A G G A C C A A A C C C A G G C T T T G C A C T T G C T A G G C A A G C	3683

```
; Patent No. 6284485
; GENERAL INFORMATION:
; APPLICANT: Boyle, William J.
; APPLICANT: Lacey, David L.
; APPLICANT: Calzone, Frank J.
; APPLICANT: Chang, Ming-Shi
; TITLE OF INVENTION: OSTEOPROTEGERIN
; NUMBER OF SEQUENCES: 53
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amgen Inc.
; STREET: 1840 Behavilland Drive
; CITY: Thousand Oaks
; STATE: California
; COUNTRY: USA
; ZIP: 91320-1789
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/795,445A
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/577,788
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Winter, Robert B.
; REFERENCE/DOCKET NUMBER: A-378
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2432 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 124..1326
; US-08-795-445A-1

Query Match 1.3%; Score 69.6; DB 3; Length 2432;
Best Local Similarity 64.0%; Pred. No. 4.8e-08;
Matches 105; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

Qy 3613 TTTTCTCCCTCTTTTTCCTCCAGAGCTGAGGACCAACCCAGGCGCTTTGCACCTT 3672
Db 1672 TTTATTTATTTTATTTTTCCTCCAGAGCTGAGGACCAACCCAGGCGCTTTGCACCTT 1731
Qy 3673 GCTAGGCAAGCGCTCTACCACTGAGCTAAATCCCAACCCAGGCGCTTTTCTTTT 3732
Db 1732 GCGAGGCAAGTGTCTTACCACTGAGCTAAATCTCAACCCGCTGAAGGCGCTTTCTTCT 1791
Qy 3733 TTTAAGACATGGTCTTATATAGTCTAGGCTGGCGCTTAAACTCA 3776
Db 1792 GCCTCTGATAGTCTATGACATCTTTTCTACAAATTCGTATCA 1835

RESULT 13
US-08-795-447A-1
; Sequence 1, Application US/08795447A
; Patent No. 6284728
; GENERAL INFORMATION:
; APPLICANT: Boyle, William J.
; APPLICANT: Lacey, David L.
; APPLICANT: Calzone, Frank J.
; APPLICANT: Chang, Ming-Shi
; TITLE OF INVENTION: Osteoprotegerin
; NUMBER OF SEQUENCES: 53
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amgen Inc.
; STREET: One Amgen Center Drive

; Patent No. 6284485
; GENERAL INFORMATION:
; APPLICANT: Boyle, William J.
; APPLICANT: Lacey, David L.
; APPLICANT: Calzone, Frank J.
; APPLICANT: Chang, Ming-Shi
; TITLE OF INVENTION: OSTEOPROTEGERIN
; NUMBER OF SEQUENCES: 53
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amgen Inc.
; STREET: 1840 Behavilland Drive
; CITY: Thousand Oaks
; STATE: California
; COUNTRY: USA
; ZIP: 91320-1789
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/795,447A
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/795,447A
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Winter, Robert B.
; REFERENCE/DOCKET NUMBER: A-378D2
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2432 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 124..1326
; US-08-795-447A-1

Query Match 1.3%; Score 69.6; DB 3; Length 2432;
Best Local Similarity 64.0%; Pred. No. 4.8e-08;
Matches 105; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

Qy 3613 TTTTCTCCCTCTTTTTCCTCCAGAGCTGAGGACCAACCCAGGCGCTTTGCACCTT 3672
Db 1672 TTTATTTATTTTATTTTTCCTCCAGAGCTGAGGACCAACCCAGGCGCTTTGCACCTT 1731
Qy 3673 GCTAGGCAAGCGCTCTACCACTGAGCTAAATCCCAACCCAGGCGCTTTTCTTTT 3732
Db 1732 GCGAGGCAAGTGTCTTACCACTGAGCTAAATCTCAACCCGCTGAAGGCGCTTTCTTCT 1791
Qy 3733 TTTAAGACATGGTCTTATATAGTCTAGGCTGGCGCTTAAACTCA 3776
Db 1792 GCCTCTGATAGTCTATGACATCTTTTCTACAAATTCGTATCA 1835

RESULT 14
US-08-974-186-1
; Sequence 1, Application US/08974186
; Patent No. 6284740
; GENERAL INFORMATION:
; APPLICANT: Boyle, William J.
; APPLICANT: Lacey, David L.
; APPLICANT: Calzone, Frank J.
; APPLICANT: Chang, Ming-Shi
; TITLE OF INVENTION: OSTEOPROTEGERIN
; NUMBER OF SEQUENCES: 53
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amgen Inc.
; STREET: 1840 Behavilland Drive
; CITY: Thousand Oaks
; STATE: California
; COUNTRY: USA
; ZIP: 91320-1789
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/974,186
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
```

Search completed: September 16, 2005, 09:17:08  
Job time : 827.974 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 15, 2005, 15:05:39 ; Search time 3234.64 Seconds  
(without alignments)  
11262.759 Million cell updates/sec

Title: US-10-650-482-3

Perfect score: 5468

Sequence: 1 cggctccctcgtcgcctg.....attaaaaaaaaaaaaaa 5468

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 7389322 seqs, 3331285599 residues

Total number of hits satisfying chosen parameters: 14778644

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : Published Applications NA:\*

1:	/cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq*
2:	/cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq*
3:	/cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq*
4:	/cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq*
5:	/cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq*
6:	/cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq*
7:	/cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq*
8:	/cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq*
9:	/cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq*
10:	/cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq*
11:	/cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq*
12:	/cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq*
13:	/cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq*
14:	/cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq*
15:	/cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq*
16:	/cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq*
17:	/cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq*
18:	/cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq*
19:	/cgn2_6/ptodata/1/pubpna/US10G_PUBCOMB.seq*
20:	/cgn2_6/ptodata/1/pubpna/US10H_PUBCOMB.seq*
21:	/cgn2_6/ptodata/1/pubpna/US10I_PUBCOMB.seq*
22:	/cgn2_6/ptodata/1/pubpna/US10J_PUBCOMB.seq*
23:	/cgn2_6/ptodata/1/pubpna/US11A_PUBCOMB.seq*
24:	/cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq*
25:	/cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq*
26:	/cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5468	100.0	5468	19 US-10-650-482-3	Sequence 3, Appli
2	1324	24.2	2942	19 US-10-650-482-1	Sequence 1, Appli
3	1189.2	21.7	2324	9 US-09-925-299-168	Sequence 168, App
4	1189.2	21.7	2324	10 US-09-925-299-168	Sequence 168, App
5	542.6	9.9	883	18 US-10-256-115-224	Sequence 224, App
6	215.8	3.9	465	22 US-10-756-149-3997	Sequence 3997, Ap
7	181.6	3.3	540	9 US-09-917-800A-1296	Sequence 1296, Ap

8	170.8	3.1	399	9 US-09-983-965-2096	Sequence 2096, Ap
9	122	2.2	242	9 US-09-783-590-12337	Sequence 12337, A
10	109	2.0	421	10 US-09-918-995-36363	Sequence 36363, A
11	99.6	1.8	1028	9 US-09-925-298-148	Sequence 148, App
12	99.6	1.8	1028	14 US-10-102-806-148	Sequence 148, App
13	98	1.8	2148	17 US-10-170-385-162	Sequence 162, App
14	98	1.8	2148	17 US-10-131-827-8849	Sequence 8849, Ap
15	79.4	1.5	8351	9 US-09-917-800A-1399	Sequence 1399, Ap
16	79.4	1.5	8351	17 US-10-388-934-665	Sequence 665, App
17	79.4	1.5	8351	17 US-10-388-934-719	Sequence 719, App
18	79.4	1.5	8351	17 US-10-191-803-1094	Sequence 1094, Ap
19	79.4	1.5	8351	18 US-10-152-319A-1461	Sequence 1461, Ap
20	79.4	1.5	119414	13 US-10-087-192-1807	Sequence 1807, Ap
21	79.2	1.4	1448	18 US-10-152-319A-2090	Sequence 2090, Ap
22	79.2	1.4	4529	17 US-10-191-803-273	Sequence 273, App
23	79.2	1.4	4529	18 US-10-152-319A-2052	Sequence 2052, Ap
24	79.2	1.4	25891	13 US-10-087-192-1657	Sequence 1657, Ap
25	79.2	1.4	53963	18 US-10-052-482-43	Sequence 43, Appl
26	78.6	1.4	1496	10 US-09-866-050A-519	Sequence 519, App
27	78.6	1.4	1496	14 US-10-152-661-519	Sequence 519, App
28	78.2	1.4	71132	13 US-10-087-192-1867	Sequence 1867, Ap
29	77.6	1.4	5531	17 US-10-191-803-234	Sequence 234, App
30	77.6	1.4	5531	18 US-10-152-319A-1974	Sequence 1974, Ap
31	77.6	1.4	215980	10 US-09-972-546-16	Sequence 16, Appl
32	77.6	1.4	215980	21 US-10-735-256-16	Sequence 16, Appl
33	77.2	1.4	233060	13 US-10-087-192-97	Sequence 97, Appl
34	77	1.4	36119	19 US-10-322-281-829	Sequence 829, App
35	76.6	1.4	6251	9 US-09-917-800A-472	Sequence 472, App
36	76.6	1.4	6251	18 US-10-152-319A-527	Sequence 527, App
37	76	1.4	479	17 US-10-388-934-727	Sequence 727, App
38	76	1.4	552	21 US-10-764-420-317	Sequence 317, App
39	76	1.4	575	21 US-10-486-708-410	Sequence 410, App
40	76	1.4	120239	19 US-10-322-281-271	Sequence 271, App
41	75.4	1.4	508	9 US-09-917-800A-319	Sequence 319, App
42	75.4	1.4	21688	21 US-10-621-911A-9	Sequence 9, Appl1
43	75.4	1.4	21688	21 US-10-646-390A-9	Sequence 9, Appl1
44	75.2	1.4	2070	16 US-10-327-813-5	Sequence 5, Appl1
45	75.2	1.4	2070	17 US-10-329-668-5	Sequence 5, Appl1
46	75.2	1.4	2070	18 US-10-152-319A-1756	Sequence 1756, Ap
47	75.2	1.4	2070	19 US-10-717-049-5	Sequence 5, Appl1
48	75.2	1.4	2070	21 US-10-899-458-13	Sequence 13, Appl
49	75	1.4	2472	17 US-10-408-713-5	Sequence 5, Appl1
50	75	1.4	2485	18 US-10-152-319A-1901	Sequence 1901, Ap
51	74.8	1.4	2010	9 US-09-809-545A-52	Sequence 52, Appl
52	74.6	1.4	227	10 US-09-930-213-739	Sequence 739, App
53	74.4	1.4	13553	17 US-10-388-934-208	Sequence 208, App
54	74.2	1.4	13011	17 US-10-388-934-36	Sequence 36, Appl
55	74.2	1.4	13011	18 US-10-152-319A-2150	Sequence 2150, Ap
56	74	1.4	3134	10 US-09-866-050A-478	Sequence 478, App
57	74	1.4	3134	14 US-10-152-661-478	Sequence 478, App
58	73	1.3	109565	19 US-10-322-281-27	Sequence 27, Appl
59	72.8	1.3	1089	20 US-10-466-634-3	Sequence 3, Appl1
60	72.8	1.3	1686	17 US-10-388-934-549	Sequence 549, App
61	72.8	1.3	2066	18 US-10-152-319A-1553	Sequence 1553, Ap
62	72.8	1.3	3561	21 US-10-486-708-210	Sequence 210, App
63	72.8	1.3	4964	17 US-10-388-934-565	Sequence 565, App
64	72.6	1.3	175338	13 US-10-087-192-1861	Sequence 1861, Ap
65	72.4	1.3	864	18 US-10-240-240A-63	Sequence 63, Appl
66	72.4	1.3	4039	17 US-10-336-091-31	Sequence 31, Appl
67	72.4	1.3	189158	13 US-10-087-192-415	Sequence 415, App
68	72.4	1.3	192673	19 US-10-331-053-1	Sequence 1, Appl1
69	72.2	1.3	23713	21 US-10-936-273-13	Sequence 13, Appl
70	72.2	1.3	32963	13 US-10-087-192-1801	Sequence 1801, Ap
71	72.2	1.3	107304	17 US-10-322-281-249	Sequence 249, App
72	72	1.3	159611	17 US-10-085-117-241	Sequence 241, App
73	71.8	1.3	983	18 US-10-240-240A-85	Sequence 85, Appl
74	71.8	1.3	1314	18 US-10-152-319A-1487	Sequence 1487, Ap
75	71.8	1.3	1314	21 US-10-486-706-374	Sequence 374, App
76	71.8	1.3	7656	21 US-10-621-911A-7	Sequence 7, Appl1
77	71.8	1.3	7656	21 US-10-646-390A-7	Sequence 7, Appl1
78	71.8	1.3	12460	14 US-10-175-523-93	Sequence 93, Appl
79	71.8	1.3	12460	21 US-11-099-268-93	Sequence 93, Appl
80	71.8	1.3	15316	21 US-10-889-101-118	Sequence 118, App





Qy	1441	GGATGGACCCACAA	CATTGCA	CAGATAAAC	CCGACGACAGGCGGTGT	CCCCCTGCTGCAGACA	1500
Db	1441	GGATGGACCCACAA	CAATTGCACAGATAA	CCGACGACAGGCGGTGT	CCCCCTGCTGCAGACA	1500	
Qy	1501	GGCCGGAGCCCACT	GAGAAAAA	CCAGAAAT	TGGTGATTCAAG	AAAGTTTTCACAGAGCCCC	1560
Db	1501	GGCCGGAGCCCACT	GAGAAAAA	CCAGAAAT	TGGTGATTCAAG	AAAGTTTTCACAGAGCCCC	1560
Qy	1561	AGGGAACGAGTCT	GTGTTTTGTGAA	TTACCCGTGGAAAA	GAAGTAATGTGAAGAGCA	CCACACTA	1620
Db	1561	AGGGAACGAGTCT	GTGTTTTGTGAA	TTACCCGTGGAAAA	GAAGTAATGTGAAGAGCA	CCACACTA	1620
Qy	1621	ATGCAACTGACCT	CTCAGATAGAGGAGAG	CCCTTCTCTGTTCTA	CCAGACCAAGTTTGTA		1680
Db	1621	ATGCAACTGACCT	CTCAGATAGAGGAGAG	CCCTTCTCTGTTCTA	CCAGACCAAGTTTGTA		1680
Qy	1681	GCAACAACTGAT	AGATTATATTTGGAGGCG	CCCCCAGTGACTT	GTGAAGCCAGGCTCTG		1740
Db	1681	GCAACAACTGAT	AGATTATATTTGGAGGCG	CCCCCAGTGACTT	GTGAAGCCAGGCTCTG		1740
Qy	1741	ATTCTGAAAGTGAG	GAATTTGGGGCGAGAA	CCCTGAGGACGAT	TGGTATTGATAGCGATGGCT		1800
Db	1741	ATTCTGAAAGTGAG	GAATTTGGGGCGAGAA	CCCTGAGGACGAT	TGGTATTGATAGCGATGGCT		1800
Qy	1801	CCCTGTCTGAAT	CAGACGTGGACAG	ACTCGGAGCCCTT	CACCTTTGGAAC	CTCTTTCC	1860
Db	1801	CCCTGTCTGAAT	CAGACGTGGACAG	ACTCGGAGCCCTT	CACCTTTGGAAC	CTCTTTCC	1860
Qy	1861	ACAGTGTAGAT	TCTTTACAAACCCCAAA	CTTTACAGCCACGAT	TACAGCGCTGCCAGAA		1920
Db	1861	ACAGTGTAGAT	TCTTTACAAACCCCAAA	CTTTACAGCCACGAT	TACAGCGCTGCCAGAA		1920
Qy	1921	TTGCCCCCAGAG	ACCCATCAGATTCAGGGA	CATCTGGTCTGGCAGCT	TGTGTAGGGA		1980
Db	1921	TTGCCCCCAGAG	ACCCATCAGATTCAGGGA	CATCTGGTCTGGCAGCT	TGTGTAGGGA		1980
Qy	1981	GCTGTCCAGGAGG	ACCCCTTCCGAGAC	CCCCGACCATAGT	TCCGGGGAGGAAGATGACT		2040
Db	1981	GCTGTCCAGGAGG	ACCCCTTCCGAGAC	CCCCGACCATAGT	TCCGGGGAGGAAGATGACT		2040
Qy	2041	GGGAACCCGAGT	GACAGTGAACAGAGAA	CTTAAATTTGTGGA	ACTCTTTCTGTCA	TTCTCTG	2100
Db	2041	GGGAACCCGAGT	GACAGTGAACAGAGAA	CTTAAATTTGTGGA	ACTCTTTCTGTCA	TTCTCTG	2100
Qy	2101	AGGAACCCCTA	CAACCTTTTAAATTTTAA	AGGCTCTTTTCA	ACCGTCAAGGAAGAA	TTGGA	2160
Db	2101	AGGAACCCCTA	CAACCTTTTAAATTTTAA	AGGCTCTTTTCA	ACCGTCAAGGAAGAA	TTGGA	2160
Qy	2161	AAGGCCGTGAGG	CTCAAGGCTCTTCTCAGGT	TCACAGTGGCTTCTCTCGGCCATCATA			2220
Db	2161	AAGGCCGTGAGG	CTCAAGGCTCTTCTCAGGT	TCACAGTGGCTTCTCTCGGCCATCATA			2220
Qy	2221	CTTTACTTTTCT	TGTAAGGCCCAGCTGTTAGAGG	CCAAAGATAATTTGTCCAGGCTGTG			2280
Db	2221	CTTTACTTTTCT	TGTAAGGCCCAGCTGTTAGAGG	CCAAAGATAATTTGTCCAGGCTGTG			2280
Qy	2281	GGCTGGGTGAGG	CTCTTCTGGAGAAAGATA	CAACCATATCAAGAGAAA	AAAGGTAA	CCCT	2340
Db	2281	GGCTGGGTGAGG	CTCTTCTGGAGAAAGATA	CAACCATATCAAGAGAAA	AAAGGTAA	CCCT	2340
Qy	2341	TCCTGGAAGAG	TTTACTGAGTATTATATA	TAGTGGTGATGAGGAT	TCGCAAGGACCATGGG		2400
Db	2341	TCCTGGAAGAG	TTTACTGAGTATTATATA	TAGTGGTGATGAGGAT	TCGCAAGGACCATGGG		2400
Qy	2401	AAGAATTTGCA	AGGGATGGATG	CGAGTTTCAGAAACGAA	TTTCAAGAAACAGAA	AGTTTGCCA	2460
Db	2401	AAGAATTTGCA	AGGGATGGATG	CGAGTTTCAGAAACGAA	TTTCAAGAAACAGAA	AGTTTGCCA	2460
Qy	2461	TTGCTACTGCT	TGGCCCTTTGAGCACA	GAGAAAAA	ATTTGTTTATATAG	CTGAGGATCGAGT	2520
Db	2461	TTGCTACTGCT	TGGCCCTTTGAGCACA	GAGAAAAA	ATTTGTTTATATAG	CTGAGGATCGAGT	2520

Qy	2521	CAAAAGGACTTACTGTGTGTACAGCAATGTTTAAAGATGGAACAGCCTCGACCAACCCGTCGCCAC	2581
Db	2521	CAAAAGGACTTACTGTGTGTACAGCAATGTTTAAAGATGGAACAGCCTCGACCAACCCGTCGCCAC	2581
Qy	2581	TCGTGCTCTTACTTTGAGAGTTTCCTTTAAAAAACAACACCTGGCAGCTGTCCTCTGGACATG	2640
Db	2581	TCGTGCTCTTACTTTGAGAGTTTCCTTTAAAAAACAACACCTGGCAGCTGTCCTCTGGACATG	2640
Qy	2641	TTTTTAAAGAAAACAACCTTGTATCTAGAGATGCGAGTTTGATTATTTTTTGGGTAAATGTGTCT	2700
Db	2641	TTTTTAAAGAAAACAACCTTGTATCTAGAGATGCGAGTTTGATTATTTTTTGGGTAAATGTGTCT	2700
Qy	2701	CATTAGAAAACACCAATCCCGATTAATGAAGAAATCTCTTATCTGTAAATCTCTCTTTTCCCTA	2760
Db	2701	CATTAGAAAACACCAATCCCGATTAATGAAGAAATCTCTTATCTGTAAATCTCTCTTTTCCCTA	2760
Qy	2761	TTTTAGTTGGATGTGGGTTTTGTCTTTTGTAGAGGGTCTCACTGCATAAATCTGTGTTGGCC	2820
Db	2761	TTTTAGTTGGATGTGGGTTTTGTCTTTTGTAGAGGGTCTCACTGCATAAATCTGTGTTGGCC	2820
Qy	2821	TAGGTCCTATGTATGTAGACCAAGGTGTGGCTTTGAAGTTGGTATTTTCCCTGTGCTCTGTGTTCC	2880
Db	2821	TAGGTCCTATGTATGTAGACCAAGGTGTGGCTTTGAAGTTGGTATTTTCCCTGTGCTCTGTGTTCC	2880
Qy	2881	TGTGCCACCAATGCCCAGCTGGAAAGTGTTTTTTAAATATTCCTCTGCTTACCTGGGGTGAGAG	2940
Db	2881	TGTGCCACCAATGCCCAGCTGGAAAGTGTTTTTTAAATATTCCTCTGCTTACCTGGGGTGAGAG	2940
Qy	2941	TTTTAAATTTTCACCTTTCAAAAACGTTTCTCTGGAGCGCAGGGCTGGTCTCGTTTTACATAC	3000
Db	2941	TTTTAAATTTTGACCTTTCAAAAACGTTTCTCTGGAGCGCAGGGCTGGTCTCGTTTTACATAC	3000
Qy	3001	AGGCTTCAGGCCAGCTAGGGCTATGTGGTGAGACCTAGTCTTAGAGGACAAAACAGAAACAA	3060
Db	3001	AGGCTTCAGGCCAGCTAGGGCTATGTGGTGAGACCTAGTCTTAGAGGACAAAACAGAAACAA	3060
Qy	3061	AACAGTCAGGTTACTGTGGAAAACCTGAGGCAAGGAGATAGGAAGGTTCAGGTATGTCCTGGA	3120
Db	3061	AACAGTCAGGTTACTGTGGAAAACCTGAGGCAAGGAGATAGGAAGGTTCAGGTATGTCCTGGA	3120
Qy	3121	ACTCAGTGAACTCTGAACTCAAGGCCAGCATGGGAGTGTAGCAAGACCTTTTTTACACTC	3180
Db	3121	ACTCAGTGAACTCTGAACTCAAGGCCAGCATGGGAGTGTAGCAAGACCTTTTTTACACTC	3180
Qy	3181	AGAATGAAAAGGAAGCTGGGGATATAGCTCGGTAGCAGAGGCCCTGCGCTACATGTGCA	3240
Db	3181	AGAATGAAAAGGAAGCTGGGGATATAGCTCGGTAGCAGAGGCCCTGCGCTACATGTGCA	3240
Qy	3241	AGGTCCTGGGTTCAAGTCCCCAGTACTGCAAAATAGAAAAGAAAACAATGCTTGGATAAC	3300
Db	3241	AGGTCCTGGGTTCAAGTCCCCAGTACTGCAAAATAGAAAAGAAAACAATGCTTGGATAAC	3300
Qy	3301	TATAAGGTTTAAAGCCTCATAGTCAGTTCTAACTCAAAATATGCATGCAATGAGTTCCTGT	3360
Db	3301	TATAAGGTTTAAAGCCTCATAGTCAGTTCTAACTCAAAATATGCATGCAATGAGTTCCTGT	3360
Qy	3361	GTGCTTTTTCTGTTCTAAAAATTAATGGGCTTTATGGGTTTTGTTTTTGTATGTTTTTATG	3420
Db	3361	GTGCTTTTTCTGTTCTAAAAATTAATGGGCTTTATGGGTTTTGTTTTTGTATGTTTTTATG	3420
Qy	3421	TCGTGTATATTTGAGGTAGGGGATTTGAGGAGGTCCTACTGTGGCCTTGAGCTCATTG	3480
Db	3421	TCGTGTATATTTGAGGTAGGGGATTTGAGGAGGTCCTACTGTGGCCTTGAGCTCATTG	3480
Qy	3481	CAGTTTTTCCGTCTCAACTGAGTGTGTTTTGTTTTGTTGTGTTTTGTTTTTACTTTCA	3540
Db	3481	CAGTTTTTCCGTCTCAACTGAGTGTGTTTTGTTTTGTTGTGTTTTGTTTTTACTTTCA	3540
Qy	3541	TAGATTTGACTTAAATGAAGGCAAAAACCTGTATCAACCTTAAAGACATTTGATGTGTCCAC	3600
Db	3541	TAGATTTGACTTAAATGAAGGCAAAAACCTGTATCAACCTTAAAGACATTTGATGTGTCCAC	3600
Qy	3601	TTCAGTGGTGGAATTTTCTCCCTCTTTTTTTTTTTTTTTTTTTTTTTTTTCCCCAGAGCTGAGGACCAAAACCCAG	3660



; TYPE: DNA		24.28; Score 1324; DB 19; Length 2942;	
; ORGANISM: Homo sapiens		Best Local Similarity 72.58; Pred. No. 0;	
US-10-650-482-1		Matches 1964; Conservative 0; Mismatches 660; Indels 84; Gaps 16;	
Qy	Db	151	CCGTGCGCGCGCGGAGGAGGCTCTCTCTATGTGTGGAGCATCTCACACGCGCTAGG 210
		96	CCATGCGCGAGCGCGTGGGGGACTCTCTATCCATGTGTGTAAGCGTCGAGCGGACTAGG 155
Qy	Db	211	ACGTCTCTCTCCCTAGCCGGATGGACCTAACCGCGGTGCGCACCGCTTCGCGGGGCTC 270
		156	GAACCTCTCTCCC-CGCGAGGATGGAAGTCGCATCAGTCGCGCGCTATTATGGCGGGGTG 214
Qy	Db	271	TGGGCGGTGCGGTGACACACTCGTGTGGGAAGCGCGCTCTCTGGGCGCTCTCTGCGGG 330
		215	TCCTCCCTGTGTCTGCGCGCGCTGCGCATTCGCTGCGCTCTGTGGCTTTCTGTGCTG 274
Qy	Db	331	CGCGGGAATCGGACTGCAGTACCCACTCCGTGGCTGGGCAAGCGGAGACTGTGTAGACC 390
		275	CTCGAAGATCGGCTTGAGCAGCGACGCCACCGCTGGGCAAGCGGAGACTCTGTAGGCT 334
Qy	Db	391	TCGGATCCAGCTGCGCTGACCGCGCTGAGCTCTGTCTCTCTCTCTCTGTCTGAGAACCGCC 450
		335	TCCTCCGAATCCCGTCGACCTCCAGCGCGCTGAGCGCGCGCGCCCTACCTGAGAGACTGTC 394
Qy	Db	451	AAG-GAAGGAGATGGAGACAGGAACGCCACGAGGCGCGGAGCGGCTGCGCTCGGCTG 509
		395	AAGAAAAGAGATGGAGCGGAGACAGGGGATCGCGGAAACGGCTTGCGCCCTCGGGCG 454
Qy	Db	510	GGCTCTCTGGTTCCGGCTGCCCC---TTCTCTTCGGCGATCGCACGCGCTCTCTCTCGGAGTTC 566
		455	GGCTTCGGTTCTGGCCACCCTTTTCTCTCGGCGATCGCAAGCAGGCTCTCTTAAGTTC 514
Qy	Db	567	CCGCGCGCTTCTCTCTGACAAAATCCCGGAACTCC-----GCTCTGCCCC 611
		515	CCGACGCTCTTGCGCCCGGAAACTCCGGGAAACCCACACTGTCTTCTCTGCGCCAGCCC 574
Qy	Db	612	GAGCGTCGGACAGGTACTGGACCAAAATTCGTTTCTCAGCTCTCTGCGCTGCTCCCTCCTAGC 671
		575	GAGACTCGGGTCAGTTACTGACAAAATCTCTCTCCAGCTCTTTCGCGCGCTCTCCCGGA 634
Qy	Db	672	CTATTCCAGAGCTGTGCTTTTGAGCCAGCTTTTCGGGGCGCTGATTTCTCTACCAATGG 731
		635	TTGCTTCAGAAAGTGTCTAAATTTGGAGCCAATTTTCGGTGGATGTTTCCGACCAAGTGG 694
Qy	Db	732	CTAGATTTGCGCAAGTTTACAGCGCTGAGAGCTTCGAGAGGACGGGAGGAATCTGAC 791
		695	CTAGATTTGCTGGAGTCTACAGCGCTGAGAGCCCTGAGAGCCCTGAGGAGCGGAGAAACCAAGCC 754
Qy	Db	792	GCTCCCAAGGTGAGAGTCTCTGAGTTACACTCGGCTGGACTCTTCGGAGAGACTCGC 851
		755	GCCCCACAGCGCAAGNAATCTTTGAGTT-CGCTGCACTCGACTCTCC---AGACCCCTC 810
Qy	Db	852	GTCTGAGTACTCTCTGCAATGGCTAGAGGAGGACTCCAGTGGCAGTGTCTGCTCTCAGA 911
		811	GGTCAACAGTCCCCTTGNATTGGCTAGAGGAGGAGATCCACTGGCAATACTCGCCCCAGA 870
Qy	Db	912	-CTGGAAGTTAACTCAAGGCCCAAGGAAGCTTTTAACTCTGACGTGCGAGCGCCACTTTCT 970
		871	CCTAAATTTGAGCTTAAGGCCAAGGGAAGTGTCTTGGACCTCGACACAGGCTTTCT 930
Qy	Db	971	CCTGAGCAGCAGCTGTGGGAGTGGAGTTGCTGCCCAAGTAGCCCTTCAAGCTGTCTAGT 1030
		931	CTTAGAGCAGCAGCTGTGGGAGTGGAGCTGTGTGCCCAAGTAGCCCTTCAATCCCGTCTGTA 990
Qy	Db	1031	CTCCACCGGAATCTGACTCTTCACTCTGCGCTCTGAGCGTTCAAGACTCTGAGACTTAGGTAA 1090
		991	CTCTAACCGGNACTTGGCTCTTCGCGCTCTGGGCGCTCTTAACATTCACGCATAGACAA 1050
Qy	Db	1091	TTTCAAGGTAGTTTCTCTATCTCTGAAACCCCTTCTCTACCTGGACTACCTTCCCCAGTTAGG 1150
Qy	Db	1051	TTTCAGTGTGGTATCCTATTTGCTGAACCCCTTCTCTACCTGGACTGCTTCTCTAGGCTAGA 1110
		1151	GCTGCGCTGTCAAGCAGCGCTGGAGGTGGCCAGTTTGTGGGTTTCCGAACACTTAACCCCC 1210
Qy	Db	1111	AGTCAGCTATCAGAACAGTGTAGGAATAGCGAGGTAGTCGGCTTCCAGACACTTAACCCC 1170
		1211	AGAGAGCTGCTATCTTTCTGAAGATGGTTGTACCCCTCAGCCGTTGCGGGCAGAGATGTC 1270
Qy	Db	1171	AGAGAGCAGCTGCTCAGAGAGGACCATTTGTATCCCCAGCGCTGAGTGCAGAACTCAT 1230
		1271	GGCAACCGCTGGAGAAGGTGTCCGCTCTCTCTACAGAGGCTGCCGGAATCCACCA 1330
Qy	Db	1231	TCGGGCTCTGTGGCAGGGATGTCCACTCTTCTCTAGAGGCTCTACAGAAATTCACCA 1290
		1331	CTTCTGTATGAACCGCTAGAAATTCCT---CCAGGCTAAACAAAGGCAAGAGTTACCCAC 1387
Qy	Db	1291	TCCTCCATGAACCGCTGGATTCTTCAACAGGCTAACAGGGGCAAGATTTACCCAC 1350
		1388	CCCTGACCAAGATAATATGGCTATCATAGCTTGAGGAGGAACTAACTTCTCCGGATGGA 1447
Qy	Db	1351	CCCTGACCAAGATAATATGGCTTACCAAGCTTGAGGAGGAAACACAGCTTCTCCGGATGGA 1410
		1448	CCCAACAATTGSCACAGATAACCCAGCAGAGCGGTGTCCCCTGCTGCAGACAGGCGC--- 1504
Qy	Db	1411	TCAAAACACTGCAGAGATAACCAACACAGTTTGTTCCTGCTGCTGGAGACATTCCTGG 1470
		1505	-----GGAGCCCACTGAGAAAAAACCCAGAAATTTGTTGATTCAAGAAAGTTTTCAC--- 1551
Qy	Db	1471	AAACACCCAGGAATCCACTGAGAAAAAATAGAAATTTAATCTACAGAGTTTCCACTTGC 1530
		1552	-----AGAGCCCCCAGGAGAGCTGTGTTTGTGAATTTACCCTGGAAAAAGA 1600
Qy	Db	1531	TTTGGAGAAGAGAGAGCCCTCTGAGGCTGTCCATCTAGTGAGATACCTATATGAAAAAGGA 1590
		1601	ATGTGAAGAGGACCACTAATGCAACTGACCTCTCAGATAGAGGAGAGAGCTTCTGT 1660
Qy	Db	1591	GCCTGGAGAGGGCGGAATAGTGTAGTTACTCATCTAGAAAGGTGACCTTCCCAT 1650
		1661	TTCTACACAGCACTGTTGTAGCAACAAACTGATAGATATATTTTGGAGCGCCCCCAG 1720
Qy	Db	1651	TTCTGCCAGNCCAGCTGTAGTAACAACTGATAGATATATTTTGGAGGTGCATCCAG 1710
		1721	TGACTTTGGAAGCCAGCTCTGATTTCTGAAAGTGAAGATTGGGGCGAGAACTCTGAGGACGA 1780
Qy	Db	1711	TGACCTGGAAAACAAGTTCTGATCCAGAAGTGAAGATTGGGATGAGGAAGCTGAGGATGA 1770
		1781	TGCTTTTGATAGCGATGGCTCCCTGTCTGNAATCAGACGTGGAACAGACTCGGAAGGCT 1840
Qy	Db	1771	TGGTTTGTAGTAGTAGTCACTGTCCAGACTCAGACCTTGAACAAAGACCTTGAAGGGCT 1830
		1841	TCACCTTTGGAACTCTTTTCCACAGTGTAGATCCTTACAAACCCCAAACTTTTACAGCCAC 1900
Qy	Db	1831	TCACCTTTGGAACTCTTTCTGAGTGTAGATCCTTATATCCCGAGAACTTTACAGCAAC 1890
		1901	GATTTCAGCGGCTGCCAGAAATTCGCCCCAGAGACCCATCAGATTCAGGGACATCTCTGGTC 1960
Qy	Db	1891	AATTCAGACTGTCTGCCAGAAATTTCTCTGAAGAGCTTCTGATTCAGAGAAGGATTTGTC 1950
		1961	TGCGAGCTGTGTGTAGGGA---GCTGTGAGGAGGAGACCCCTTCCCGAGAGACCCCGACCA 2017
Qy	Db	1951	TGGCAAGTCTGATCTAGAGAAATTCCTCCAGCTTGGAAAGCTTCTCTGAGACCCCTGAGCA 2010
		2018	TAGTTCGGGAGGAGATGACTGGGAACCGAGTGCAGATGGAAGCAGAGAACTCTTAAAT 2077
Qy	Db	2011	TAGTTCGGGAGGAGAGATGACTGGGAATCTAGTGCAGATGGAAGCAGAGAGTCTCAAACT 2070
		2078	GTGGAACCTCTTCTGTCTCATTTCTGAGGACCCCTACAACTTTTAAATTTTAAAGGCTCCTTT 2137
Qy	Db	2071	GTGGAACCTCATTTCTGTAAATTTCTGATGACCCCTACAACTTTTAAATTTTAAAGGCTCCTTT 2130
		2138	TCAACCGTCAAGGAGAAATTTGGAAGGCGGTGAGGACTCAAGGCTCTTCTGAGGTAC 2197

Db 2131 TCAAAACATCAGGGGAAATGAGAAAGGCTGTGCTGACTCAAAAGACCCCATCTGAGTCCAT 2190  
 Qy 2198 AGTGGCCCTCTCTGGCCATCATACCTTACTTCTTTGTAAGGCCAGCTGTTAGAGAGCCA 2257  
 Db 2191 TGTGGCCCATTTCTGAGTGTCAACACTTACTTCTTTGTAAGGTGACGCTGTTGGGGAGCCA 2250  
 Qy 2258 AGAAGATAAATGCTCCAGGCTGTGGGCTGGGTGAGGCTCTTGTGTGAGAAAGATACACCCA 2317  
 Db 2251 AGAAGATGNAATGTCCAGACTCGGTACAGGCTGACGCTTCTTCTGGAGGAGACACACACA 2310  
 Qy 2318 TATCAAGAGAAAAAGTAACTCTCTCGAAGAGTTACTGAGTATTATATAAGTGTGGA 2377  
 Db 2311 TGTCAAAAGAAAAAGTAACTCTCTCGAAGAGTTACTGAGTATTATATAAGTGTGGA 2370  
 Qy 2378 TGAGATCGCAAGACCAATGCGGAGAAATTTGCAAGGGATGGATCGAGTTCAGAAACG 2437  
 Db 2371 TGAGATCGCAAGACCAATGCGGAGAAATTTGCAAGGGATGGATCGAGTTCAGAAACG 2430  
 Qy 2438 AATTCAAGAAAACAGAAAGTTGCCATTTGGCTACTGCTTGGCCCTTTGAGCACAGAGAAAAAT 2497  
 Db 2431 AATTCAAGAAAACAGAAAGTGTCTTGGATATTGCTTGACATTTGAACACAGAGAAAGAT 2490  
 Qy 2498 GTTTAATAGACT---GAGGATCGAGTCAAAAGGACTTACTGTGTGACAGCAATGTTAAGA 2553  
 Db 2491 GTTTAATAGACTCCAGGGAAACATGCTTCAAAAGGACTTAATGTCTCAAGCAATGTTGAGT 2550  
 Qy 2554 AGTGA-ACAGCTGCAACCGGTGCCACTCTGCTCTTACTTACTGAGA-GTTTCCCTTAAAA 2611  
 Db 2551 TGGCAGCCTGTAGTCTTAGCTAGCATACACTACCTCTTACCTGAGAGGTGCTTTTAAAA 2610  
 Qy 2612 ACAAAACACTGGCAGCTGTCTTGGACATGTTTTAAAGAAACAACTTGTATCTAGAGATG 2671  
 Db 2611 ACAAACTTGGCAGCTGTCTTGGACATGTTTTAAAGAAACAACTTGTATCTAGAGATG 2670  
 Qy 2672 CAGTTGATATTTTTGGGTAAATGTCATATAGAA-----ACACCAACT 2717  
 Db 2671 AGTTTAAATTTTTTTTTTGGCAACATATCCCACTCAGAAACATTCAGGTTTGAAGCCAGCC 2730  
 Qy 2718 CCGATAATGAGAAAT-----CTCTATCTGTAATCCCTCTCTTTCTTATTTAGTTGGAT 2771  
 Db 2731 CTGATAATGAGGATGAACCTAGTGTGATTTCTAATCCTCCCTTTTGTGATTTAGTTGGAT 2790  
 Qy 2772 GTGGGTTT 2779  
 Db 2791 GTGCTTTT 2798

RESULT 3

US-09-925-299-168  
 ; Sequence 168, Application US/09925299  
 ; Patent No. US20020055627A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Rosen et al.  
 ; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
 ; FILE REFERENCE: P102  
 ; CURRENT APPLICATION NUMBER: US/09/925,299  
 ; CURRENT FILING DATE: 2001-08-10  
 ; PRIOR APPLICATION NUMBER: PCT/US00/05883  
 ; PRIOR FILING DATE: 2000-03-08  
 ; PRIOR APPLICATION NUMBER: 60/124,270  
 ; PRIOR FILING DATE: 1999-03-12  
 ; NUMBER OF SEQ ID NOS: 1556  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 168  
 ; LENGTH: 2324  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 US-09-925-299-168

Query Match 21.7%; Score 1189.2; DB 9; Length 2324;  
 Best Local Similarity 75.3%; Pred. No. 0;  
 Matches 1642; Conservative 4; Mismatches 477; Indels 59; Gaps 11;

Qy 519 TTCCGGCTGCCCTTCTTTCGGCGATCGACCGCTGCTCTTTCGGAGTTTCCCGCCGCTTCC 578  
 Db 43 TTTTCAACCTCTCTGCTAGTTCGGGATCGAAGCAGGCTCTTCTAAGTTTCCGACGCTCTT 102  
 Qy 579 TCTCGACAAATATCCGGGAACTCCGCTCT-----GCCCGAGCGTGGAGCC 623  
 Db 103 GGCCCGGAAAAATCCGGGAAACCCACACATGCTTCTCTGCCCCAGCGCGAGCTCGGGTC 162  
 Qy 624 AGGTACTGGACCAATTTGCTTCTCAGCTCTTTCGCTGCTGCTCCCTAGCTATTTCCAGAG 683  
 Db 163 AGTTACTGGACCAATTTGCTTCTCAGCTCTTTCGCTGCTGCTCCCGGATTTGCTTCAAG 222  
 Qy 684 CTGCTGCTTTTGGAGCCAGCTTTTCGGGGGCTGATTTCTTACCAAGATGGCTAGATTTTGGC 743  
 Db 223 GTGCTAATTTTGGAGCCAACTTTTCGGTGGAAATTTTCGGACCAAGATGGCTAGATTTTGGT 282  
 Qy 744 GCAAGTTACAGCGCCCTGAGAGCTTCGAGAGAGCGGAGGAATCTGACGCTCCACCGTG 803  
 Db 283 GGAGTCTACAGCGCCCTGAGAGCCCTGAAGGAGCGGAGAAACCCAGCGCGCCCCACAGCG 342  
 Qy 804 CAGAAGTCTCTGAGTTACACTCGGCGTGGACTCTTTCGCGAAGACTCGCGTCTGCTCAGTACT 863  
 Db 343 CAGAAATCTTTGAGTT--CGCTGACGCTCGACTCTCTC---AGACCTCTCGGTCAACAGTCC 398  
 Qy 864 CTTGCAATTTGGCTAGAGAGGAGACTCCAGTGGCAGTGTCTCGTCTCTCAGA-CTGGAAAGTTAA 922  
 Db 399 CTTGATTTGGCTAGAGAGGAGGATCCACTGGCAATACTCGCCCCCAGACCTTAAATTTGGA 458  
 Qy 923 ACTCAAGGCCCAGAAAGAGCTTTAGACTCTGACGCGCCCACTTTTCTCTCTGGAGCAGCA 982  
 Db 459 GCTTAAGGCCAAAGGAAAGTGTCTTGGACCTTGCAGCACAAGGCTTTTCTCTTAGAGCAGCA 518  
 Qy 983 GCTGTGGGAGTGGAGTTGCTGCCAGTGTCTTCAAGCTGTCTAGTCTCTCCCAACCCGAGA 1042  
 Db 519 GCTGTGGGAGTGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 578  
 Qy 1043 ACTTGAATCTTCAATCTCTGCGGCTCTGAGCGTTTCAAGCTTTAGGTAAATTTCAAGTGTAGT 1102  
 Db 579 ACTTGGCTCTTCGCGCTCTGCGCTTCTTAAACATTTCAACGCAATAGACAATTTCAAGTGTGT 638  
 Qy 1103 TTCTATCTCTGAAACCTTCTTACTGAGACTACTTCCCGAGTTAGGCTGCGCTGTCA 1162  
 Db 639 ATCTATTTGCTGAACCTTCTTACTACCTGGAATGCTTTCTTCTAGGCTAGAAAGTCAAGTATCA 698  
 Qy 1163 GAGCAGCGCTGGAGTGGCCAGTTTGTGGTGTTCGGAACACTAACCCACAGAGACTGCTA 1222  
 Db 699 GAACAGTGTAGAAATAGCAGAGTGTGCTGCTTCCAGACTAACCCACAGAGCAGCTG 758  
 Qy 1223 TCTTTCTGAAGATGTTGTCACTCTGAGCGTTGCGGGCAGAGATGTCGGCAACCGCTG 1282  
 Db 759 CTTGAGAGAGGACCATTTGTCATCCCGAGCGGTGARTGCAGAACTCATTTCCGSGCTCGTG 818  
 Qy 1283 GAGAGGTGTCCGCTCTCTTACAGAAAGGCTGCGGAAATCCACCACTTCTGATGAA 1342  
 Db 819 GCAGGGATGTCCACTCTTCTTACGGAAGGCTTACCAAGAAATTCACCATCTTCCGATGAA 878  
 Qy 1343 ACGGCTAGAAATTCCT---CCAGGCTTAAAGGAGGAGTGTACCCACCTCTGACCAAGA 1399  
 Db 879 ACGGCTGAAATTCCTTCAACAGGCTAGCAAGGGGCAAGATWTACCCACCTCTGACCAAGA 938  
 Qy 1400 TAATGGCTATCATAGCTGGAGGAGAAATTAACCTTCTCCGGATGGACCCCAACATTTG 1459  
 Db 939 TAATGGCTTACCACAGCTGGAGGAGGAAACACAGCTTCTCCGGATGGATCCAAACACTG 998  
 Qy 1460 CACAGATAACCCAGCAGCGGCTGCTCCCTGCTGCGAGACAGGCC-----GGA 1507  
 Db 999 CAGAGATAACCCAAACACAGATTTGTTCTGCTGCTGCGAGACATTCCTCTGGAACACCCAGGA 1058  
 Qy 1508 GCCCACTGAGAAAAAACAGAAATTTGGTGTATTAAGAAAGTTTTCAC-----A 1552  
 Db 1059 ATCCACTGAGAAAAAATAGAAATTAATTAATACAGAGGTTCCACTTCTTTTGGAGAGA 1118  
 Qy 1553 GAGCCCCCAGGGAAGCAGTCTGTTTGTGAAATTAACCGTGGAAAAAAGAAATGTGAGAGGA 1612

[illegible]

```

Db          2199  AGCTGTCTTTGACATTTTTT 2220

RESULT 4
US-09-925-299-168
; Sequence 168, Application US/09925299
; Publication No. US20030040617A9
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 168
; LENGTH: 2324
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-299-168

```

Query Match	21.7%	Score 1189.2	DB 10	Length 2324	
Best Local Similarity	75.3%	Pred. No. 0			
Matches 1642	Conservative	4	Mismatches 477	Indels 59	Gaps 11
Qy	519	TTCCGGCTGCGCCCTTCCTTCGGCGCATCGCAGCGTGTCTTCGAGATTCCCGCGCCTTCC	578		
Db	43	TTTCAACCCCTCTGTCTAGTCGGCGATCGCAAGCAGGCTCTTCTAAGTTCCTCGACGCTCTT	102		
Qy	579	TCTCGACAAAATCCCGGAATCCCGTCTCT-----GCCGAGCGTCGGACC	623		
Db	103	GGCCCGGAAACTCCGGGAACCCCACTGCTTCTCTGCCCAGCCCGAGACTCGGTC	162		
Qy	624	AGTACTGGACCAAAATTCCTTTCTAGCTCCTTTCGCCCTGCTCCCTAGCTATTCCAGAAG	683		
Db	163	AGTTACTGGACGAAACTGCTCTCCAGCTCCTTTGCGCGCTCCCGGATTGCTTCAGAA	222		
Qy	684	CTGCTGCTTTGGAGCCAGCTTTCCGGGGGCTGATTCCTACCATAGTGCTAGATTTCGCC	743		
Db	223	GTGCTAAATTTGGAGCCCACTTTTCGGTGGAAATGTTTTCCGACCATAGTGCTAGATTTCGT	282		
Qy	744	GCAAGTTTACAGCCCTTGAGAGCTTCGAGAGGACGGAGGAAATCGACGCTCCACCGTG	803		
Db	283	GGAGTCTACAGCGCCTGAGAGCCCTGAAGGAGACGGAGAAACAGCCGCCCCCAACAGCG	342		
Qy	804	CAGAAGTCTCTGAGTTACACTGCGGGCTGGACTCTTCGCGAAGACTCGCGCTCGTCAGTACT	863		
Db	343	CAGAAATCTTTAGTT- CGTGCAGCTCGACTCCTC---AGACCCCTCGGTCAACAGTCC	398		
Qy	864	CTTGCAATGGCTTAGAGAGGGAATCAAGTGGCAGTGCTCGTCTCTCAGA-CTGGAAAGTTAA	922		
Db	399	CCTTGAATGGCTTAGAGAGGGGATCCACTGGCAATATCTGCCCCCAGACCTAAATTTGGA	458		
Qy	923	ACTCAGGCCCCAGGAAGAGCTTTTAGACTCTGCAGGCGCCACCTTCTCTCTGGAGCAGCA	982		
Db	459	GCTTAAAGGCCCAAGGGAAGTGCTTTTGAACCTCGCAGACACAGGCTTTTCTCTTAGAGCAGCA	518		
Qy	983	GCTGTGGGAGTGGAGTTGCTGCGCAGTAGCTTTCAAGCTGTGTTAGTCTCCACCACGAGA	1042		
Db	519	GCTGTGGGAGTGGAGCTGTTGCCCCAGTAGCTTCAATCCCGTCTGTACTCTTAACCGGGA	578		
Qy	1043	ACTTGACTCTTTCATCTCTTGGGCTCTGAGCGTTCAGAGCTTAGGTAAATTTCAAGTAGT	1102		
Db	579	ACTTGGCTCTTCGCCCTCTGGGCTTCTTAAACATTTCAACGCATAGACAAATTTTCAGTGTGGT	638		
Qy	1103	TTGCTATCTCCTGAAACCTTCTCCTACCTGGACTACTTCCCCAGTTAGGGCTGCGCTGTCA	1162		
Db	639	ATCCTATTGTGCTGAAACCTTCTCTACTCTGACTGCTTTTCTTAGCTAGAACTAGCTATCA	698		







```
Qy 1959 TCTGCACTGTGTGTTAGGA---GCTGTGAGAGAGGACCCCTTCCGGAGACCCCGAC 2015
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
241 TCTGGCAAGTCTGATCTAGAGAAATCTCCAGTCTGGAAGCTTCTCTGAGACCCCTGAG 300
Qy 2016 CATAGTTCCGGGAGGAGATGACTGGGAAACGAGTGCAGATGAAGCAGAGAACTTAAA 2075
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
301 CATAGTTCTGGGAGGAGATGACTGGGAAATCTAGTGCAGATGAAGCAGAGATCTCAA 360
Qy 2076 ---TTGTGGAATCTTCTGTCA---TTCTGAGGACCCCTCAACCTTTTAAATTTTAA 2129
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
361 ACTGTGGGAATCTTCAATCTGTAAATCTGATGACCCCTCAACCCCTTTAAATTTTAA 420
Qy 2130 GCTCTTTTCAACCGTCAAGGAGAAATTTGGAAGCCGTCAAGACTCAAGGCTCTTCT 2189
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
421 GCTCTTTTCAACATCAAGGGGAAATGAGAAAGGCTGTCTGAGACTCAAGACCCCTCT 480
Qy 2190 GAGGTCAAGTGGCTTCTCTGGCCATCATACCTTTACTTCTTGTAAAGCCAGCTGTTA 2249
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
481 GAGTCCATTGTGGCCATTTCTGAGTGTCAACCTTACTTCTTGTAAAGTGCAGCTGTT 540
Qy 2250 GAGAGCCAAAGATAATTTGTCAGGCTGTGGGCTGGGTGAGGCTTCTGCTGGAGAA 2309
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
541 GGGAGCCAAAGAGTGAATGTCCAGACTCGGTACAGGTGAGCTTCTTCTGGAGGA 600
Qy 2310 TACACCCATATCAAGAGAAAAAGTAACTTCTCTGGAAGAGTACTGAGTATATATA 2359
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
601 CACACATGTGTCAAGAAAAAAGTAACTTCTCTGAAAGAGTACTGAGTATATATA 660
Qy 2370 AGTGTGATGAGGATCGCAAGGACCATGGGAAGATTTGCAAGGATGATGAGGTT 2429
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
661 AGTGTGATGAGGATCGCAAGGACCATGGGAAGATTTGCAAGGATGATGAGGTT 720
Qy 2430 CAGAAACGAATTCAGAAACAGAAAGTTGCCATTTGGCTACTCTTGGCTTTTGACACAGA 2489
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
721 CAGAAACGAATTCAGAAACAGAAAGTCTATTGGATATCTTGACATTTGAACACAGA 780
Qy 2490 GAAAAATGTTTATAGACT---GAGATCGATCAAGAGACTTACTGTTGTACAGCAA 2545
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
781 GAAAGATGTTTATAGACTCCAGGAACATGCTTCAAGGACTTAATGTTCTCAAGCAA 840
Qy 2546 TGTTAAGAGTGAAC 2560
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
841 TGTGAGTTGCAGC 855

RESULT 6
US-10-756-149-3997
; Sequence 3997, Application US/10756149
; Publication No. US20050181375A1
; GENERAL INFORMATION:
; APPLICANT: Zlotnik, Albert
; TITLE OF INVENTION: NOVEL METHODS OF DIAGNOSIS OF METASTATIC CANCER, COMPOSITIONS AND
; TITLE OF INVENTION: METHODS OF SCREENING FOR MODULATORS OF METASTATIC CANCER
; FILE REFERENCE: file
; CURRENT APPLICATION NUMBER: US/10/756.149
; CURRENT FILING DATE: 2004-01-12
; NUMBER OF SEQ ID NOS: 5818
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 3997
; LENGTH: 465
; TYPE: DNA
; ORGANISM: Homo Sapiens
US-10-756-149-3997

Query Match 3.9%; Score 215.8; DB 22; Length 465;
Best Local Similarity 73.4%; Pred. No. 6e-48;
Matches 343; Conservative 0; Mismatches 117; Indels 7; Gaps 5;
```

```
Qy 758 CTTGAGAGCTTCGAGAGACGGGAGGAATCTGACGCTCCACGCTGCAGAGAGTCTCTGAG 817
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1 CCTGAGAGCCCTGAGGAGCGGAGAACCAAGCGCCGCCCAAGGAGAAATCTTTGAG 60
```

```
Qy 818 TTACACTGCGCGTGGACTCTTTCGGGAAGACTCGCGTCTGCTCAGTACTCTTTCGATTTGGCTAG 877
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
61 TT-CGCTGCAGCTCGAATCTCCTC---AGACCCCTCGGTCAACAGTCCCTCTTGAATGGCTAG 116
Qy 878 AGAGAGGACTCCAGTGGCAGTGTCTGTCCTCAGA-CTGGAAGTTTAAACTCAAGGCCCAGG 936
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
117 AGAGAGGATCCACTGGCAATACTTCGCCCCACAGCTTAAATTTGGAGCTTAAAGCCCAAGG 176
Qy 937 AAAGAGCTTTAGACTCTGCAGAGCCCACTTTTCTCTCGAGCAGCAGCTGTGGGGAGTGG 996
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
177 GAAGTGTCTTTGGACCTTCGACAGCAGAGCTTTTCTTTAGAGCAGCAGCTGTGGGGAGTGG 236
Qy 997 AGTTGTGCCAGTAGAGCTTCAAGCTGTCTAGTCTCTCCACCGAGAACTTGAATCTTCAT 1056
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
237 AGCTGTTGCCAGTAGCTTTCAATCCGCTCTGACTCTAAACCGGAACTTTGGCTCTTCGC 296
Qy 1057 CCTCTGGGCTCTGAGGTTTCAGAGCTTAGGTAATTTCAAGGTAGTTTCTCTATCTCCTGA 1116
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
297 CCTCTGGGCTTTAAACATTTCAACGATAGCAATTTCACTGTGTGTATCTTATTTGCTGA 356
Qy 1117 ACCCTTCTACTCTGGACTACTTCCCGAGTTAGGGCTGCGCTGCTCAG-AGCAGCGCTGGA 1175
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
357 ACCCTTCTACTCTGGACTGCTTTTCTAGGCTAGAGTCACTATCAGAAACAGTATGGA 416
Qy 1176 GTGGGCCAGTTTGTGGGTTTCCGAAACAC-TAACCCCGCAGAGAGCTGCT 1221
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
417 AATAGCGAGGTAGTCTGGCTTCCAGACACTTAAACCCCGCAGAGAGCT 463
```

```
RESULT 7
US-09-917-800A-1296/c
; Sequence 1296, Application US/09917800A
; Patent No. US20020119462A1
; GENERAL INFORMATION:
; APPLICANT: Mendrick, Donna
; APPLICANT: Porter, Mark
; APPLICANT: Johnson, Kory
; APPLICANT: Castle, Arthur
; APPLICANT: Elashoff, Michael
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Molecular Toxicology Modeling
; FILE REFERENCE: 44921-5038-US
; CURRENT APPLICATION NUMBER: US/09/917, 800A
; CURRENT FILING DATE: 2001-07-31
; PRIOR FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: US 60/222,040
; PRIOR FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: US 60/222,880
; PRIOR FILING DATE: 2000-11-02
; PRIOR APPLICATION NUMBER: US 60/290,029
; PRIOR FILING DATE: 2001-05-11
; PRIOR APPLICATION NUMBER: US 60/290,645
; PRIOR FILING DATE: 2001-05-15
; PRIOR APPLICATION NUMBER: US 60/292,336
; PRIOR FILING DATE: 2001-05-22
; PRIOR APPLICATION NUMBER: US 60/295,798
; PRIOR FILING DATE: 2001-06-06
; PRIOR APPLICATION NUMBER: US 60/297,457
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,884
; PRIOR FILING DATE: 2001-06-19
; PRIOR APPLICATION NUMBER: US 60/303,459
; PRIOR FILING DATE: 2001-07-09
; NUMBER OF SEQ ID NOS: 1740
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1296
; LENGTH: 540
; TYPE: DNA
; ORGANISM: Rattus norvegicus
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020119462A1 AI237580
US-09-917-800A-1296
```

Query Match 3.3%; Score 181.6; DB 9; Length 540;

Best Local Similarity		75.8%;	Pred. No. 1.7e-38;
Matches 388;		Conservative 0;	Mismatches 69; Indels 55; Gaps 11;
Qy	4960	ATATTTAAAAATTTCTCCAGATTACACCTTCACCTATCAAAATGAGTAAATGAC-----	5011
Db	499	ATGTTTGAAAAATGTTCTCCAGATTATACCTTCACCTATCAAAATGAATAAATGACTGGTTAAA	440
Qy	5012	-----TCATGCTCGAGACATGCTCTGATGTGGGCAAGAACGAAGGATCT-----TTG	5060
Db	439	ATAATACGCGCTGCAGAGATGCTCAAAATGGTGGCAAGACAGAGAGGCTCTAAACTGAAGG	380
Qy	5061	ACTCAAGCAGAAAAACTGTCTATTCTCATCCAGGCCCCCAGGAAAGAACACCTCCCAAGGCA	5120
Db	379	AGAAAAACAAAATTTACTGTCTATTGTCTCAT- CGAATCCCAGGAAAGAACACCTCGAGGCA	321
Qy	5121	GGCAGGCAGGCGAGCAGGCAATGTTGGTCTCTAGTTGGAATACACATTC- AAGTCTTGCAGTG	5179
Db	320	-----AGGCAGGCAATGGCGGTCTTAGCTGAAACACTTACACTCAGTCTTGCAGTG	272
Qy	5180	GTGCTTTT- -AGATCTGTGTAGCATGTACAGGCTCTGTACAGGTGGG- GCCACACTTCTGA	5236
Db	271	GTGCTTTTAGAGATCTCTGTAGCATGTGAGGCTCTGTACAGGTGGGACCCACACTTCCGA	212
Qy	5237	GGGCTGAAATGTGGCAACCCCTTTATCTAACTTGAATCAAAAACCGTCAAAATTTTATTTT	5296
Db	211	GTGCTGAAATGTGGCAACCC- TTATCTAACTTGAATCAAAAACCGTCAAGATTTTATTTT	153
Qy	5297	TATAATTTAAGAA-----AGAGTTGGGGAATGACATTTTTTTGAGTTGGCCTTTTCAGTCA	5352
Db	152	TATAAATTTAAGAAAGTAGAGACGGAAATGAGAAATTTTTTTTTTTTGGAAATGAGATGTTT	93
Qy	5353	G-----TCAATTTTACGTGTAAC- -GTGGAGATTTTGATCTCAGACTTATATT	5397
Db	92	GAGTTGGTTCAGATCATTTTACATGTAACATAGTGTAGATTTTATATCGCGAGATTTATATT	33
Qy	5398	TGTATATAATTTATTAACATAATCTCTAAATGTT	5429
Db	32	TGTATATAATTTATTAACATAATCTCTAAATGTT	1

RESULT 8  
US-09-983-965-2096  
; Sequence 2096, Application US/09983965  
; Patent NO. US20020137160A1  
; GENERAL INFORMATION:  
; APPLICANT: Warren, Wesley C.  
; APPLICANT: Tao, Nengbing  
; APPLICANT: Byatt, John C.  
; APPLICANT: Mathialagan, Nagappan  
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND  
; TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION  
; FILE REFERENCE: 37-21(10297)C  
; CURRENT APPLICATION NUMBER: US/09/983,965  
; CURRENT FILING DATE: 2001-10-26  
; PRIOR APPLICATION NUMBER: US 09/465,231  
; PRIOR FILING DATE: 1999-12-15  
; PRIOR APPLICATION NUMBER: US 60/113,678  
; PRIOR FILING DATE: 1998-12-17  
; NUMBER OF SEQ ID NOS: 5912  
; SEQ ID NO 2096  
; LENGTH: 399  
; TYPE: DNA  
; ORGANISM: Bos taurus  
; FEATURE:  
; OTHER INFORMATION: Clone ID: 32-LIB3057-012-Q1-K1-H11  
US-09-983-965-2096

Query Match 3.1%; Score 170.8; DB 9; Length 399;  
Best Local Similarity 79.1%; Pred. No. 1.3e-35;  
Matches 231; Conservative 0; Mismatches 52; Indels 9; Gaps 2;  
Qy 2330 AAGGTTAACTTCTCGAAGACTTACTGAGTATTATATTAAGTCGTGATGAGGATCGCAA 2389

Db	102	ATAGGTTAACTTCCTTGAAGAAGTTACTAGTATTACATAAGCGGTGATGAGGATCGAAA	161
Qy	2390	AGGACCATGGGAAGAAATTTTCAAGGGATGGATCGAGGTTCCAGAAACGAATTCAGAAGAAC	2449
Db	162	AGGACCATGGGAAGAAATTTTGCACGGGATGGATGTCAGGTTCCAGAAACGAATTCAGAAGAAC	221
Qy	2450	AGAAATTGGCCATTTGGCTTACTGCTTGGCCCTTTGACGACAGAGAAATA-----ATGTTTA	2502
Db	222	AGAAGAAGCTATTGGGATCTGCTTGGACNTTTGAGCACAGAGAAATAAGATGTTTAAATATAA	281
Qy	2503	ATAGACTGAGGATCGAGTCAAAGGACTTACTGTGTGACAGCAATGTTTAAAG--AGTGAAC	2560
Db	282	CTTCAAGCAACAATGTTTCAAGAGCACTTAATGTTTCGACCAATGTTAAGATGATTCGAC	341
Qy	2561	AGCGTCGAACCCGTGGCCACTCTGTCTCTTACTTGAGAGHTTCCCTTAAAAA	2612
Db	342	AGCCTCTGGCCCTTAGATACACACTCTTACTTTGAGAGGGTACTTTTAAAAA	393

RESULT 9  
US-09-783-590-12337  
; Sequence 12337, Application US/09783590  
; Patent No. US20020110850A1  
; GENERAL INFORMATION:  
; APPLICANT: Dillon, Patrick J.  
; APPLICANT: Haseltine, William A.  
; APPLICANT: Li, Haodong  
; APPLICANT: Rosen, Craig A.  
; APPLICANT: Ruben, Steven M.  
; TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2  
; FILE REFERENCE: PO-16.2C1

```

1 FILE REFERENCE: 10:201
2 CURRENT APPLICATION NUMBER: US/09/783,590
3 CURRENT FILING DATE: 2000-02-15
4 PRIOR APPLICATION NUMBER: 08/420,856
5 PRIOR FILING DATE: 1995-04-12
6 PRIOR APPLICATION NUMBER: 08/346,731
7 PRIOR FILING DATE: 1994-11-21
8 NUMBER OF SEQ ID NOS: 12485
9 SOFTWARE: PatentIn Ver. 2.0
10 SEQ ID NO 12337
11 LENGTH: 242
12 TYPE: DNA
13 ORGANISM: Homo sapiens
14 FEATURE:
15 NAME/KEY: misc feature
16 LOCATION: (1)
17 OTHER INFORMATION: n equals a,t,g, or c
18 NAME/KEY: misc feature
19 LOCATION: (4)
20 OTHER INFORMATION: n equals a,t,g, or c
21 NAME/KEY: misc feature
22 LOCATION: (8)
23 OTHER INFORMATION: n equals a,t,g, or c
24 NAME/KEY: misc feature
25 LOCATION: (35)
26 OTHER INFORMATION: n equals a,t,g, or c
27 NAME/KEY: misc feature
28 LOCATION: (216)
29 OTHER INFORMATION: n equals a,t,g, or c
30 NAME/KEY: misc feature
31 LOCATION: (220)
32 OTHER INFORMATION: n equals a,t,g, or c
33 NAME/KEY: misc feature
34 LOCATION: (221)
35 OTHER INFORMATION: n equals a,t,g, or c
36 US-09-783-590-12337

```

Query Match 2.2%; Score 122; DB 9; Length 242;  
Best Local Similarity 82.7%; Pred. No. 2.3e-22;  
Matches 162; Conservative 0; Mismatches 32; Indels 2; Gaps 2;  
Qy 1982 CTGTCAGGAGGCCCTTCGGAGAGACCCCGACCATGTTCCGGGGAGAGATGACTG 2041

Db 22 CTCCAGTCTGGAGCCCTTCTCTGAGACCCCTGAGCAGTAGTTCTGGGGAGGAAGATGACTG 81  
 Qy 2042 GGAACCCAGTGCAGATGAAGCAGAGAGATCTTAAATTTGGAACT-CTTTCTGTCTATCTG 2100  
 Db 82 GGGATCTAGTCAGATGAAGCAGAGAGTCTCAAACTGTGGAACCTCCATCTGTAACTCTG 141  
 Qy 2101 -AGAACCCCTCAACCTTTTAAATTTTAAAGGCTCTTTTCAACCGTCAGGGAAGAAATGG 2159  
 Db 142 AATGACCCCTCAACCCCTTTAAATTTTAAAGGCTCTTTTCAACATCAGGAGGAAATGAG 201  
 Qy 2160 AAAGCGCTCAGGACT 2175  
 Db 202 AAAGCGTCTGTGNT 217

RESULT 10  
 US-09-918-995-36363  
 ; Sequence 36363, Application US/09918995  
 ; Publication No. US20030073623A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Hyseq, Inc.  
 ; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
 ; FROM VARIOUS CDNA LIBRARIES  
 ; FILE REFERENCE: 20411-756  
 ; CURRENT APPLICATION NUMBER: US/09/918,995  
 ; PRIOR FILING DATE: 2001-07-30  
 ; PRIOR FILING DATE: 2001-07-30  
 ; PRIOR FILING DATE: 1999-01-20  
 ; NUMBER OF SEQ ID NOS: 38054  
 ; SOFTWARE: FastSeq for Windows Version 3.0  
 ; SEQ ID NO 36363  
 ; LENGTH: 421  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 US-09-918-995-36363

Query Match 2.0%; Score 109; DB 10; Length 421;  
 Best Local Similarity 71.6%; Pred. No. 1.3e-18;  
 Matches 187; Conservative 0; Mismatches 65; Indels 9; Gaps 3;  
 Qy 5214 ACAGTGGGGCCACACTCTGAGGCTGAAATGTGGCAACCTTTATC--TAACTTGAA 5271  
 Db 132 ACAGGTAAGATGCCATTTCTGAGTGATGAATGTAACACTTCTTCATCTTTAACTTGA 191  
 Qy 5272 ATCAAAACCGTCAAAATTTATTTT-----TATAATTTAAGAAAGAGTTGGGAATGAC 5325  
 Db 192 ATCAAACTATCAGATTTTATTTTGTATATTTTAAAGAGTTAAAGTTAGGGACTAGA 251  
 Qy 5326 ATTTTGTAGTGGCCCTTTTCAGCTCAGTCATTT-TACGTGTAAACGTGGAGATTGTAG 5384  
 Db 252 AGACTCTAAATTTGGCTTCTACAGATCAATATTTTAAATGTAACCTAGTTGGGATTTATAG 311  
 Qy 5385 CTCAGATATATTTGTATATATTTTAACTAATCTGTAATTTGTAATTAATAATATTGTC 5444  
 Db 312 ATAAATTTATTTTGGCTGTATATACATAACTAATCTGTAATTTGTAATTAATAATATTGTC 371  
 Qy 5445 AATTATTAAAAA 5465  
 Db 372 AATTATTAAAAA 392

RESULT 11  
 US-09-925-298-148  
 ; Sequence 148, Application US/09925298  
 ; Publication No. US20020039764A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Rosen et al.  
 ; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
 ; FILE REFERENCE: P103  
 ; CURRENT APPLICATION NUMBER: US/09/925,298  
 ; PRIOR FILING DATE: 2001-08-10  
 ; PRIOR FILING DATE: 2000-03-08

; PRIOR APPLICATION NUMBER: 60/124,270  
 ; PRIOR FILING DATE: 1999-03-12  
 ; NUMBER OF SEQ ID NOS: 846  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 148  
 ; LENGTH: 1028  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 US-09-925-298-148

Query Match 1.8%; Score 99.6; DB 9; Length 1028;  
 Best Local Similarity 70.8%; Pred. No. 8.9e-16;  
 Matches 177; Conservative 0; Mismatches 64; Indels 9; Gaps 3;  
 Qy 5214 ACAGTGGGGCCACACTCTGAGGCTGAAATGTGGCAACCTTTATC--TAACTTGAA 5271  
 Db 757 ACAGGTAAGATGCCATTTCTGAGTGATGAATGTAACACTTCTTCATCTTTAACTTGA 816  
 Qy 5272 ATCAAAACCGTCAAAATTTATTTT-----TATAATTTAAGAAAGAGTTGGGAATGAC 5325  
 Db 817 ATCAAACTATCAGATTTTATTTTGTATATTTTAAAGAGTTAAAGTTAGGGACTAGA 876  
 Qy 5326 ATTTTGTAGTGGCCCTTTTCAGCTCAGTCATTT-TACGTGTAAACGTGGAGATTGTAG 5384  
 Db 877 AGACTCTAAATTTGGCTTCTACAGATCAATATTTTAAATGTAACCTAGTTGGGATTTATAG 936  
 Qy 5385 CTCAGATATATTTGTATATATTTTAACTAATCTGTAATTTGTAATTAATAATATTGTC 5444  
 Db 937 TTAATTTATTTTGGTGTATATACATAACTAATCTGTAATTTGTAATTAATAATATTGTC 996  
 Qy 5445 AATTATTAAA 5454  
 Db 997 AATTATTAAA 1006

RESULT 12  
 US-10-102-806-148  
 ; Sequence 148, Application US/10102806  
 ; Publication No. US20030054421A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Rosen et al.  
 ; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
 ; FILE REFERENCE: P103P1C1  
 ; CURRENT APPLICATION NUMBER: US/10/102,806  
 ; PRIOR FILING DATE: 2002-03-22  
 ; PRIOR FILING DATE: 2002-03-22  
 ; PRIOR FILING DATE: 2001-08-10  
 ; PRIOR APPLICATION NUMBER: PCT/US00/05981  
 ; PRIOR FILING DATE: 2000-03-08  
 ; PRIOR FILING DATE: 1999-03-12  
 ; NUMBER OF SEQ ID NOS: 846  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 148  
 ; LENGTH: 1028  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 US-10-102-806-148

Query Match 1.8%; Score 99.6; DB 14; Length 1028;  
 Best Local Similarity 70.8%; Pred. No. 8.9e-16;  
 Matches 177; Conservative 0; Mismatches 64; Indels 9; Gaps 3;  
 Qy 5214 ACAGTGGGGCCACACTCTGAGGCTGAAATGTGGCAACCTTTATC--TAACTTGAA 5271  
 Db 757 ACAGGTAAGATGCCATTTCTGAGTGATGAATGTAACACTTCTTCATCTTTAACTTGA 816  
 Qy 5272 ATCAAAACCGTCAAAATTTATTTT-----TATAATTTAAGAAAGAGTTGGGAATGAC 5325  
 Db 817 ATCAAACTATCAGATTTTATTTTGTATATTTTAAAGAGTTAAAGTTAGGGACTAGA 876  
 Qy 5326 ATTTTGTAGTGGCCCTTTTCAGCTCAGTCATTT-TACGTGTAAACGTGGAGATTGTAG 5384

Db 877 AGACTCTAAATTGGCTCTACAGATCAATAAATTTAAATGTAACTAGTTGGGATTTATAG 936  
Qy 5385 CTCAGATTATATTCTGTATATATTATTAACATACTGTAATGTAATAATGTAATAATATATTGC 5444  
Db 937 TTAAATATATATTCTGTATATATAACATACTAATCTGTAATGTAATAATGTAATAATATATTGC 996  
Qy 5445 AATTATTAAA 5454  
Db 997 AATTATTAAA 1006

## RESULT 13

US-10-170-385-162  
; Sequence 162, Application US/10170385  
; Publication No. US2003020372A1  
; GENERAL INFORMATION:  
; APPLICANT: Mundy, Neil Raymond  
; APPLICANT: Mundy, Christopher Robert  
; APPLICANT: Kan, On  
; APPLICANT: Harris, Robert Alan  
; APPLICANT: White, Jonathan  
; APPLICANT: Binley, Katie Mary  
; APPLICANT: Rayner, William Nigel  
; APPLICANT: Naylor, Stuart  
; APPLICANT: Kingsman, Susan Mary  
; APPLICANT: Krige, David  
; TITLE OF INVENTION: ANALYSIS METHOD  
; FILE REFERENCE: 53268200100  
; CURRENT APPLICATION NUMBER: US/10/170,385  
; PRIOR FILING DATE: 2002-06-12  
; PRIOR APPLICATION NUMBER: PCT/GB02/01662  
; PRIOR FILING DATE: 2002-04-08  
; PRIOR APPLICATION NUMBER: PCT/GB01/05458  
; PRIOR FILING DATE: 2001-12-10  
; NUMBER OF SEQ ID NOS: 549  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 162  
; LENGTH: 2148  
; TYPE: DNA  
; ORGANISM: Homo Sapiens  
US-10-170-385-162

Query Match 1.8%; Score 98; DB 17; Length 2148;  
Best Local Similarity 70.4%; Pred. No. 4.1e-15;  
Matches 176; Conservative 0; Mismatches 65; Indels 9; Gaps 3;  
Qy 5214 ACAGTGGGGCCACACTTCTGAGGCTGAAATGTGGCAACCCCTTATC--TAACTTGAA 5271  
Db 1888 ACAGTAAAGTTCCATTCTGAGTGAATGAAATGTAACACTTCTTTCATCTTTAACTTGAA 1947  
Qy 5272 ATCAAAACCGTCAAAATTTTATTTT-----TATAATTTAAGAAAGTGTGGGAATGAC 5325  
Db 1948 ATCAAAACTATCAGATTTTATTTTGTATATATTTAAGGAAGGTAAAGTTAGGGACTAGA 2007  
Qy 5326 ATTTTTCAGTTGGCCTTTCAGCTCAGTCATTT-TACGTGTAACGTGGAGATTGTATAG 5384  
Db 2008 AGACTCTAAATTTGGCTCTACAGATCAATAATTTAAATGTAACCTAGTTGGGATTTATAG 2067  
Qy 5385 CTCAGATTATTTGTATATATTAATTAACATACTGTAATCTGTAATAATATATATTGC 5444  
Db 2068 TTAAATATATATTCTGTATATAACATAATTAATCTGTAATGTAATAATATATATTGC 2127  
Qy 5445 AATTATTAAA 5454  
Db 2128 AATTATTAAA 2137

## RESULT 14

US-10-131-827-8849  
; Sequence 8849, Application US/10131827  
; Publication No. US20040009479A1  
; GENERAL INFORMATION:  
; APPLICANT: Wohlgemuth, Jay

; APPLICANT: Fry, Kirk  
; APPLICANT: Woodward, Robert  
; APPLICANT: Ly, Ngoc  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING AND MONITORING AUTOIMMUNE  
; TITLE OF INVENTION: CHRONIC INFLAMMATORY DISEASES  
; FILE REFERENCE: 506612000120  
; CURRENT APPLICATION NUMBER: US/10/131,827  
; CURRENT FILING DATE: 2002-09-06  
; PRIOR APPLICATION NUMBER: US 10/006,290  
; PRIOR FILING DATE: 2001-10-22  
; PRIOR APPLICATION NUMBER: US 60/296,764  
; PRIOR FILING DATE: 2001-06-08  
; NUMBER OF SEQ ID NOS: 9090  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 8849  
; LENGTH: 2148  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-131-827-8849  
Query Match 1.8%; Score 98; DB 17; Length 2148;  
Best Local Similarity 70.4%; Pred. No. 4.1e-15;  
Matches 176; Conservative 0; Mismatches 65; Indels 9; Gaps 3;  
Qy 5214 ACAGTGGGGCCACACTTCTGAGGCTGAAATGTGGCAACCCCTTATC--TAACTTGAA 5271  
Db 1888 ACAGTAAAGTTCCATTCTGAGTGAATGAAATGTAACACTTCTTTCATCTTTAACTTGAA 1947  
Qy 5272 ATCAAAACCGTCAAAATTTTATTTT-----TATAATTTAAGAAAGTGTGGGAATGAC 5325  
Db 1948 ATCAAAACTATCAGATTTTATTTTGTATATATTTAAGGAAGGTAAAGTTAGGGACTAGA 2007  
Qy 5326 ATTTTTCAGTTGGCCTTTCAGCTCAGTCATTT-TACGTGTAACGTGGAGATTGTATAG 5384  
Db 2008 AGACTCTAAATTTGGCTCTACAGATCAATAATTTAAATGTAACCTAGTTGGGATTTATAG 2067  
Qy 5385 CTCAGATTATTTGTATATATTAATTAACATACTGTAATCTGTAATAATATATATTGC 5444  
Db 2068 TTAAATATATATTCTGTATATAACATAATTAATCTGTAATGTAATAATATATATTGC 2127  
Qy 5445 AATTATTAAA 5454  
Db 2128 AATTATTAAA 2137  
RESULT 15  
US-09-917-800A-1399  
; Sequence 1399, Application US/09917800A  
; Patent No. US20020119462A1  
; GENERAL INFORMATION:  
; APPLICANT: Mendrick, Donna  
; APPLICANT: Porter, Mark  
; APPLICANT: Johnson, Kory  
; APPLICANT: Castle, Arthur  
; APPLICANT: Elashoff, Michael  
; APPLICANT: Gene Logic, Inc.  
; TITLE OF INVENTION: Molecular Toxicology Modeling  
; FILE REFERENCE: 44921-5038-US  
; CURRENT APPLICATION NUMBER: US/09/917,800A  
; CURRENT FILING DATE: 2001-07-31  
; PRIOR APPLICATION NUMBER: US 60/222,040  
; PRIOR FILING DATE: 2000-07-31  
; PRIOR APPLICATION NUMBER: US 60/222,880  
; PRIOR FILING DATE: 2000-11-02  
; PRIOR APPLICATION NUMBER: US 60/290,029  
; PRIOR FILING DATE: 2001-05-11  
; PRIOR APPLICATION NUMBER: US 60/290,645  
; PRIOR FILING DATE: 2001-05-15  
; PRIOR APPLICATION NUMBER: US 60/292,336  
; PRIOR FILING DATE: 2001-05-22  
; PRIOR APPLICATION NUMBER: US 60/295,798  
; PRIOR FILING DATE: 2001-06-06  
; PRIOR APPLICATION NUMBER: US 60/297,457

; PRIOR FILING DATE: 2001-06-13  
; PRIOR APPLICATION NUMBER: US 60/298,884  
; PRIOR FILING DATE: 2001-06-19  
; PRIOR APPLICATION NUMBER: US 60/303,459  
; PRIOR FILING DATE: 2001-07-09  
; NUMBER OF SEQ ID NOS: 1740  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1399  
; LENGTH: 8351  
; TYPE: DNA  
; ORGANISM: Rattus norvegicus  
; FEATURE:  
; OTHER INFORMATION: Genbank Accession No. US20020119462A1 M11794  
US-09-917-800A-1399

Query Match	1.5t;	Score 79.4;	DB 9;	Length 8351;
Best Local Similarity	64.8t;	Pred. No. 1.3e-09;		
Matches 138;	Conservative 0;	Mismatches 56;	Indels 9;	Gaps 1;

  

Qy	3504	GTTTGGTTTGTGTTGTTGTTGTTTACTTTTCATAGATTGACTTAATGAAGCAA	3563
Db	6264	GTTTGTGTTTTCGTTTTTTTTTTGTTTATTGTTTGTGTTTGTGTTTATCTAATGCAG	6323
Qy	3564	AAACCTGTATCAACCTAAAGACATTGATGTGTCACCTTCAGTGGTGGATTTCCTCCCTC	3623
Db	6324	TCCCTGGATATCACCTAAATGATCCCTCTGCTCGGTTTTTTTTTTTTTTTTTTTTTTT	6383
Qy	3624	TTTTTTTTTTTTTCCCC-----CAGAGCTGAGGACCAACCCAGGCTTTGCACTTGC	3674
Db	6384	TTTTTTTTTTTGGTTCTTTTTTCGGAGCTGGGACCGAACCAGGCGCTTGGCGCTCC	6443
Qy	3675	TAGCAAGCGCTCTACCACTGAGCTAAATCCCC	3707
Db	6444	TAGTAAGCGCTCTACCACTGAGCTAAATCCTC	6476

Search completed: September 16, 2005, 10:40:03  
Job time : 3240.64 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 15, 2005, 11:51:14 ; Search time 16323.4 Seconds  
(without alignments)  
12750.757 Million cell updates/sec

Title: US-10-650-482-3  
Perfect score: 5468  
Sequence: 1 cggctctccgtctgccttg.....attaaaaaaaaaaaaa 5468

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 100 summaries

Database :  
EST: \*  
1: gb\_est1: \*  
2: gb\_est2: \*  
3: gb\_hic: \*  
4: gb\_est3: \*  
5: gb\_est4: \*  
6: gb\_est5: \*  
7: gb\_est6: \*  
8: gb\_ges1: \*  
9: gb\_ges2: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2641	48.3	2731	3	AK049028 Mus muscu
2	2522.8	46.1	2619	3	AK082957 Mus muscu
3	1271	23.2	1342	3	AK086606 Mus muscu
4	828.2	15.1	871	5	BQ922494 AGENCOURT
5	810	14.8	966	5	BQ926670 AGENCOURT
6	808.6	14.8	997	1	AV055059 AV055059
7	799.2	14.6	890	7	CF583619 AGENCOURT
8	796	14.6	1011	2	BB610365 BB610365
9	794.2	14.5	913	7	CF583620 AGENCOURT
10	786.6	14.4	946	6	CA978373 AGENCOURT
11	786.4	14.4	3941	3	HSM805059
12	774.8	14.2	801	7	CO428800 UI-M-HX0
13	767.4	14.0	796	4	BI654191 603280489
14	763	14.0	854	7	CF746885 UI-M-HX0
15	759.6	13.9	835	7	CF617044 AGENCOURT
16	754.8	13.8	936	4	BI737283 AGENCOURT
17	739.8	13.5	876	5	BQ953605 AGENCOURT
18	737.4	13.5	954	5	BQ943218 AGENCOURT
19	732.8	13.4	811	6	CA750998 UI-M-HX0
20	730	13.4	744	7	CF749066 UI-M-HX0
21	720.4	13.2	725	6	CD352623 UI-M-HX0
22	717.2	13.1	784	7	CK021950 AGENCOURT
23	717	13.1	717	5	BQ444408 UI-M-HX0
24	715.8	13.1	731	7	CO426562 UI-M-HX0



|||||  
302 ACAGGGCCGGAAGGGCGCTGGCCCTCGGGCTGGGCTCTGGTTCCGGCTGCCCTTCCTTC 361  
Qy  
538 GGCGATCGCAGCGCTGCTCTTCGGAGTTCCCGCGCCCTCTCTCTCGACAAATCCCGGA 597  
Db  
362 GGCGATCGCAGCGCTGCTCTTCGGAGTTCCCGCGCCCTCTCTCTCGACAAATCCCGGA 421  
Qy  
598 ACTCCGCTCTGCCCGAGCGTCGACACAGGTACTGGACCAATTCCTTTCTCAGCTCCTTG 657  
Db  
422 ACTCCGCTCTGCCCGAGCGTCGACACAGGTACTGGACCAATTCCTTTCTCAGCTCCTTG 481  
Qy  
658 CCCTGCTCCCTAGCCTATTCAGAAGCTGCTGCTTTGGAGCCAGCTTTCCGGGGCCCTGA 717  
Db  
482 CCCTGCTCCCTAGCCTATTCAGAAGCTGCTGCTTTGGAGCCAGCTTTTCGGGGCCCTGA 541  
Qy  
718 TTCCTACCAAGATGCTAGATTTCCCGCAAGTTACAGCGCCCTCAGAGCTTCGAGAGGAC 777  
Db  
542 TTCCTACCAAGATGCTAGATTTTCGCGCAAGTTACAGCGCCCTCAGAGCTTTGAGAGGAC 601  
Qy  
778 GGGAGGAATCTGACGCTCCCAAGGTGCAGAACTCTCTGAGTTACACTGGGCTGGACTCT 837  
Db  
602 GGGAGGAATCTGCGGCTCCCAAGGTGCAGAACTCTCTGAGTT-CACTGGGCTGGACTCT 660  
Qy  
838 TCGGGAAGACTCGCGTCTGCTAGTACTCTTGGCAATTTGGCTAGAGAGGACTCCAGTGGAG 897  
Db  
661 TCGGAAGACTT--GGTTGTCAGTTCTCTTG-ATTGGCTAGAGGAGGACTGCAGTGGCAG 717  
Qy  
898 TGCTCGTCTCAGACTCGGAGTT-AACTCAAGGCCCGAGGAAGAGCTTTAGACTCTGCA 956  
Db  
718 TGCTCGTCTCAGACTCGGAGTTAAACTCAAGGCCCGAGGAAGAGCTTTAGACTCTGCA 777  
Qy  
957 GCGCCCACTTTCTCTCGGAGCAGAGCTGTGGGAGTGGAGTTGCTGCCAGTAGCCTT 1016  
Db  
778 GCGCCCACTTTCTCTCGGAGCAGAGCTGTGGGAGTGGAGTTGCTGCCAGTAGCCTT 837  
Qy  
1017 CAAGCTGCTAGTCTCCCAAGAACTTGACTCTTCACTCTCGGCTCTGAGCGTT 1076  
Db  
838 CAAGCTGCTAGTCTCCCAAGAACTTGACTCTTCACTCTCGGCTCTGAGCGTT 897  
Qy  
1077 CAGAGCTTAGGTAAATTCAGGTAGTTTCCATCTCTGAAACCTTCTACTCTGAGCTAC 1136  
Db  
898 CAGAGCTTAGGTAAATTCAGGTAGTTTCCATCTCTGAAACCTTCTACTCTGAGCTAC 957  
Qy  
1137 CTTCGCCAGTTAGGCTGCGCTGTCCAGAGCAGCGCTGGAGTGGCCAGTTTGTGGGTTTC 1196  
Db  
958 CTTCGCCAGTTAGGCTGCGCTGTCCAGAGCAGCGCTGGAGTGGCCAGTTTGTGGGTTTC 1017  
Qy  
1197 CGAACACTAACCCAGAGAGCTGCTATCTTTCTGAAAGATGTTGTCACTCAGCCGTTG 1256  
Db  
1018 CGAACACTAACCCAGAGAGCTGCTATCTTTCTGAAAGATGTTGTCACTCAGCCGTTG 1077  
Qy  
1257 CGGCGAGAGATGTCCGCAACCGCTGGAGAGGTGTCCGCTCTCTCTACAGAGGCGCTG 1316  
Db  
1078 CGGCGAGAGATGTCCGCAACCGCTGGAGAGGTGTCCGCTCTCTCTACAGAGGCGCTG 1137  
Qy  
1317 CCGGAAATCCACCACTTCGTATGAAACGGCTAGAAATTCCTCCAGGCTTAAACAAAGGGCAA 1376  
Db  
1138 CCGGAAATCCACCACTTCGTATGAAACGGCTAGAAATTCCTCCAGGCTTAAACAAAGGGCAA 1197  
Qy  
1377 GAGTTACCCACCCCTGACCAAGATAATGGCTATCATAGCTGGAGGAGGAACATAACCTT 1436  
Db  
1198 GAGTTACCCACCCCTGACCAAGATAATGGCTATCATAGCTGGAGGAGGAACATAACCTT 1257  
Qy  
1437 CTCGGATGGACCCACACATTCACAGATAAACCCAGCAGAGCGGTGTCCCTGCTGCA 1496  
Db  
1258 CTCGGATGGACCCACACATTCACAGATAAACCCAGCAGAGCGGTGTCCCTGCTGCA 1317  
Qy  
1497 GACAGGCGGAGCCCACTGAGAAAAAACCAAAATTTGGTATTCAAGAGTTTCAAGAGC 1556  
Db  
1318 GACAGGCGGAGCCCACTGAGAAAAAACCGGAAATTTGGTATTCAAGAGTTTCAAGAGC 1377  
Qy  
1557 CCCAGGGAAGAGCTGTGTTTGTGAATTAACCGGTGGAAGAAAGATGTGAAGAGGACAC 1616  
|||||

Db  
1378 CCCAGGGAAGCAGTCTGTTTTGTGTAATTACCGTGGAAAAAGAAATGTGAAGAGGACCAC 1437  
Qy  
1617 ACTAATGCAACTGACCTCTCAGATAGAGGAGAGCCTTCTCTGTTTCTACAGACCAAGTT 1676  
Db  
1438 ACTAATGCAACTGACCTCTCAGATAGAGGAGAGCCTTCTCTGTTTCTACAGACCAAGTT 1497  
Qy  
1677 TGAGCAACAAACTGATAGATTATATTTTGGGAGGCGCCCCAGTGTGACTTTGGAAGCCAGC 1736  
Db  
1498 TGAGCAACAAACTGATAGATTATATTTTGGGAGGCGCCCCAGTGTGACTTTGGAAGCCAGC 1557  
Qy  
1737 TCTGATTTCTGAATGAGGANTGGGGCGAGGAACCTGAGGACGATGGCTTTGTATAGCGAT 1796  
Db  
1558 TCTGATTTCTGAATGAGGANTGGGGCGAGGAACCTGAGGACGATGGCTTTGTATAGCGAT 1617  
Qy  
1797 GGCTCCCTGCTCGAATCAGACGTCGAAACAGGACTCGGAAGGCTTCACTTTTGGAACTCT 1856  
Db  
1618 GGCTCCCTGCTCGAATCAGACGTCGAAACAGGACTCGGAAGGCTTCACTTTTGGAACTCT 1677  
Qy  
1857 TTCCACAGTGTAGATCTTTACAAAAACCCCAAAACTTTTACGCCACGATTCAGACGGCTGCC 1916  
Db  
1678 TTCCACAGTGTAGATCTTTACAAAAACCCCAAAACTTTTACGCCACGATTCAGACGGCTGCC 1737  
Qy  
1917 AGAATTCGCCCCCAGAGACCCATCAGATTCAGGACATCTCTGCTTGGCAGCTGTGGTGA 1976  
Db  
1738 AGAATTCGCCCCCAGAGACCCATCAGATTCAGGACATCTCTGCTTGGCAGCTGTGGTGA 1797  
Qy  
1977 GCGAGCTGTCCAGAGGAGCCCTTCCGGAGACCCCGACCATAGTTCCGGGAGGAGAGAT 2036  
Db  
1798 GCGAGCTGTCCAGAGGAGCCCTTCCGGAGACCCCGACCATAGTTCCGGGAGGAGAGAT 1857  
Qy  
2037 GACTGGGAACCCGAGTCAGATGAAGCAGAGAACTTTAAATTTGTGGAACTCTTTCTGTCTCAT 2096  
Db  
1858 GACTGGGAACCCGAGTCAGATGAAGCAGAGAACTTTAAATTTGTGGAACTCTTTCTGTCTCAT 1917  
Qy  
2097 TCTGAGGACCCCTTACAACTTTTAAATTTTAAAGCTCTCTTCAACCGTCAGGGAAGAAAT 2156  
Db  
1918 TCTGAGGACCCCTTACAACTTTTAAATTTTAAAGCTCTCTTCAACCGTCAGGGAAGAAAT 1977  
Qy  
2157 TGGAAAGGCCGTCCAGGACTCAAAAGGCTCTTCTGAGGTACAGTGGGCTTCTCTGGCCAT 2216  
Db  
1978 TGGAAAGGCCGTCCAGGACTCAAAAGGCTCTTCTGAGGTACAGTGGGCTTCTCTGGCCAT 2037  
Qy  
2217 CATACCTTACTTTCTTTGTAAGGCCAGCTTTAGAGGCCAAAGAAATATTTGTCCAGGC 2276  
Db  
2038 CATACCTTACTTTCTTTGTAAGGCCAGCTTTAGAGGCCAAAGAAATATTTGTCCAGGC 2097  
Qy  
2277 TGTGGGCTGGGTGAGGCTCTTGTGGAGAAAGATAACCCATATCAGAGAAAAAAGGTA 2336  
Db  
2098 TGTGGGCTGGGTGAGGCTCTTGTGGAGAAAGATAACCCATATCAGAGAAAAAAGGTA 2157  
Qy  
2337 ACCTTCTCGAAGAAAGTTACTGTAGTATTTATTAAGTGGTGTATGAGGATCGCAAGAGGACCA 2396  
Db  
2158 ACCTTCTCGAAGAAAGTTACTGTAGTATTTATTAAGTGGTGTATGAGGATCGCAAGAGGACCA 2217  
Qy  
2397 TGGGAAGAAATTTGCAAGGGAATGATGAGGTTCCAGAAAAAGAAATTTCAAGAAAAACAGAGTT 2456  
Db  
2218 TGGGAAGAAATTTGCAAGGGAATGATGAGGTTCCAGAAAAAGAAATTTCAAGAAAAACAGAGTT 2277  
Qy  
2457 GCCATTTGGCTACTGCTTGGGCTTTGAGCAGAGAAAAAATGTTTATAGACTGAGGATC 2516  
Db  
2278 GCCATTTGGCTACTGCTTGGGCTTTGAGCAGAGAAAAAATGTTTATAGACTGAGGATC 2337  
Qy  
2517 GAGTCAAAAGGACTTACTGTTGTACAGCAATGTTTAAAGAGTGAACAGGCTCGCAACCCGTCG 2576  
Db  
2338 GAGTCAAAAGGACTTACTGTTGTACAGCAATGTTTAAAGAGTGAACAGGCTCGCAACCCGTCG 2397  
Qy  
2577 CCACCTGCTCTTCTTACTGTGAGGTTTCCCTTAAAAAACAAACACTGGCAGCTGTCTTGGGA 2636  
Db  
2398 CCACCTGCTCTTCTTACTGTGAGGTTTCCCTTAAAAAACAAACACTGGCAGCTGTCTTGGGA 2457  
Qy  
2637 CATGTTTTTAAAGAAACAACTTGTATCTAGAGATGCAGTTTGTATTTTGGGTATGT 2696  
Db  
2458 CATGTTTTTAAAGAAACAACTTGTATCTAGAGATGCAGTTTGTATTTTGGGTATATGT 2517  
|||||

```
QY 2697 GTCCTATTAGAAACCAACTCCGATATCGAATCTCTTATCTGTATGTAATCTCTCTTTT 2756
|||||
Db 2518 GTCCTATTAGAAACCAACTCCGATATCGAATCTCTTATCTGTATGTAATCTCTCTTTT 2577
|||||
QY 2757 CCTATTATTAGTGGATGTGGTTTCTCTTTTGTAGAGGCTCTCACTGATTAATCTGTTT 2816
Db 2578 CCTATTATTAGTGGATGTGGTTTCTCTTTTGTAGAGGCTCTCACTGATTAATCTGTTT 2637
|||||
QY 2817 GGCCTAGCTCTATGTATGTAGACCAAGTGGCTTGAAGTTGGTATTTTCTTGCTCTGT 2876
|||||
Db 2638 GGCCTAGCTCTATGTATGTAGACCAAGTGGCTTGAAGTTGGTATTTTCTTGCTCTGT 2697
|||||
QY 2877 TTCTCTGTGCCACCATGCCAGCT 2899
Db 2698 TTCTCTGTGCCACCATGCCAGCT 2720
|||||

RESULT 2
AK082957
LOCUS
DEFINITION
Mus musculus 12 days embryo spinal cord cDNA, RIKEN full-length
enriched library, clone: C530022L24 product: hypothetical protein,
full insert sequence.
ACCESSION
AK082957
VERSION
AK082957.1 GI:26350136
KEYWORDS
HTC; CAP trapper.
SOURCE
Mus musculus (house mouse)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1
Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
99279253
10349636
2
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
20499374
11042159
3
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multipipillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
20530913
11076861
4
The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)
5
The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
6 (bases 1 to 2619)
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,
Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,
```

```

Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,
Okazaki, Y., Saito, K., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akashira, S.,
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
Muramatsu, M. and Hayashizaki, Y.
Direct Submission
Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.jp,
URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222,
Fax: 81-45-503-9216)
cdna library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Please visit our web site for further details.
URL: http://genome.gsc.riken.jp/
URL: http://fantom.gsc.riken.jp/.
FEATURES
Location/Qualifiers
1..2619
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM DB:C530022L24"
/db_xref="taxon:10090"
/clone="C530022L24"
/tissue_type="spinal cord"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="12 days embryo"
418..2511
CDS
(note="unnamed protein product; hypothetical protein
putative")
/codon_start=1
/protein_id="BAC38708.1"
/db_xref="GI:26350137"
/translation="METGTHRRKRPGRPRGSRWFLPFLRRSHACSEFPFPPSRQNP
GNSALPERRTRYWTKLLSOLLPSLFOKLWLSQFGLIPTWLDLFAASYSALRA
LRGRESAAPTOKSLSSLRLEDSSDLVSSLDWLEGLWCQSSDLKLELKAORRA
LDSAAPTFLLEQLWGVLELLPSSIQAGLVSHRELDSSSGPLSVQSLGNFKVSYLLN
PSYLDYLPQGLRCQSSAGGQFVGFRITLTPESCILSEDCGHQFPLKRAESATAWRKC
PPLSTGLPEIHLRMKRLLEFLQANKGQELPTPDQNGYHSLSEEHLLRMDPQHCTD
NPAQAVSPADRPETPEKPELVIEVVSQSSPLFCPLVKECEDEEDHTNATDLS
RGESLPVSTRPVCNKLIIDYILGASDLEASDSSEDMGPEPDDGPDGDSLS
RQDSLEGHLHNSFHSVDYPNQFTATITQARTAPRDPDSGTSWSGCGVSGCQ
EGELPTPDSSEEDDMEPSADEANLKNLWNSFCHSESDYLLNFKAPQPQSGKNWK
GRQDSKASSEATVAFSGHITLLSKAQLLSDNCPGCGLGEALAGERTHTLRKKV
TFLERVTEYISGDEDRKGFWEFARDGCRFKRIQIETVAIGYCLAFEHREKMFNL
RIESKDLILLYSNVKK"
ORIGIN
Query Match 46.1%; Score 2522.8; DB 3; Length 2619;
Best Local Similarity 98.9%; Pred. No. 0;
Matches 2593; Conservative 0; Mismatches 22; Indels 7; Gaps 5;
QY 46 CATTTTGAGCTTCGCTCTTTTGGCCCTCGCCTGCCACCCAGCCACCTTTTCGCCCTTGG 105
|||||
Db 2 CATTTTGAGCTTCGCTCTTTTGGCCCTCGCCTGCCACCCAGCCACCTTTTCGCCCTTGG 61
|||||
QY 106 CGTTTCGCGCTTCGTCGCGGCCACCGGAAAACGCGCCGTCGTCGTCGTCGCGCGCGCGC 165
|||||
Db 62 CGTTTCGCGCTTCGTCGCGGCCACCGGAAAACGCGCCGTCGTCGTCGTCGCGCGCGCGC 121
|||||
QY 166 GAGGAGGGCTTCTCTATGTCGGCGATCTCACAGGCTAGGAGCTCTCTTCCTTCCCTA 225
|||||
Db 122 GAGGAGGGCTTCTCTATGTCGGCGATCTCACAGGCTAGGAGCTCTCTTCCTTCCCTA 181
|||||
```

Qy	226	GC	GGGATGGACCTAACCGCGGTGCCCAACCGCTTGC	CGGCGCTCTGGCGCGTCCCGTGC	285
Db	182	GC	CGGATGGACCTAACCGCGGTGCCCAACCGCTTGC	CGGCGCTCTGGCGCGTCCCGTGC	241
Qy	286	AG	CACCTGTTGGGAAAGCGCGCTCTCTGGGCTCTCT	GGCGCTCTGGCGCGGGAATCGGACT	345
Db	242	AG	CACCTGTTGGGAAAGCGCGCTCTCTGGGCTCTCT	GGCGCTCTGGCGCGGGAATCGGACT	301
Qy	346	GC	AGTACCACTCCGCTGGCTGGGCAAGCGGAGACTGT	GAGACTGTGAGACTCGGATCCAGCTGC	405
Db	302	GC	AGTACCACTCCGCTGGCTGGGCAAGCGGAGACTGT	GAGACTGTGAGACTCGGATCCAGCTGC	361
Qy	406	GCT	AGCGCGCTGAGCTCTGCTCTCTCTGCTGTGAGAG	CGCGCAAGGAAAGAGATGG	465
Db	362	GCT	AGCGCGCTGAGCTCTGCTCTCTCTGCTGTGAGAG	CGCGCAAGGAAAGAGATGG	421
Qy	466	AG	CAGGAAACGACAGGCGCGGAGCGGCTGCGCTGCG	CTGGGCTCTGGTTCGGGC	525
Db	422	AG	CAGGAAACGACAGGCGCGGAGCGGCTGCGCTGCG	CTGGGCTCTGGTTCGGGC	481
Qy	526	TG	CCCTCTCTTCGCGGATCGACGCCCTGCTCTCGGAG	TTCCCGCGCTTCCCTCTCGAC	585
Db	482	TG	CCCTCTCTTCGCGGATCGACGCCCTGCTCTCGGAG	TTCCCGCGCTTCCCTCTCGAC	541
Qy	586	AAAA	TCCCGGAACTCCGCTCTGCCGAGCGTCGGACCA	GGTACTGGACCAAAATTCCTTT	645
Db	542	AAAA	TCCCGGAACTCCGCTCTGCCGAGCGTCGGACCA	GGTACTGGACCAAAATTCCTTT	601
Qy	646	CT	CAGCTCTTCCCTGCTCCCTAGCCTATTTCCAGAG	CTGCTCTTGGAGCCAGCTTT	705
Db	602	CT	CAGCTCTTCCCTGCTCCCTAGCCTATTTCCAGAG	CTGCTCTTGGAGCCAGCTTT	661
Qy	706	CC	GGGCGCTGATTTCTACAGATGGCTAGATTTTGC	GCAAGTTACAGGCCCTGAGAG	765
Db	662	TC	GGGCGCTGATTTCTACAGATGGCTAGATTTTGC	GCAAGTTACAGGCCCTGAGAG	721
Qy	766	CTT	CGAGAGGACGGGAGAACTGACCGTCCCGGTGCA	AGTCTCTGAGTTACACTG	825
Db	722	CTT	CGAGAGGACGGGAGAACTGACCGTCCCGGTGCA	AGTCTCTGAGTTACACTG	780
Qy	826	CG	CTGAGCTCTTCCGAGAGACTCGGCTGCTCAGTA	CTCTTGCAATTCCTAGAGAGGA	885
Db	781	CG	CTGAGCTCTTCCGAGAGACTCGGCTGCTCAGTA	CTCTTGCAATTCCTAGAGAGGA	837
Qy	886	CT	CAGTGGCAGTGTCTGCTCTCAGA - CTGGAAGT	TAAACTCAAGGCCCAGGAAAGACT	944
Db	838	CT	CAGTGGCAGTGTCTGCTCTCAGA - CTGGAAGT	TAAACTCAAGGCCCAGGAAAGACT	897
Qy	945	TT	AGACTCTGAGCGGCCACTTTCTCTCTGAGCAG	CAGCTGTGGGAGTGGAGTTGCTG	1004
Db	898	TT	AGACTCTGAGCGGCCACTTTCTCTCTGAGCAG	CAGCTGTGGGAGTGGAGTTGCTG	957
Qy	1005	CC	CAGTAGCTTCAAGCTGCTAGTCTCCACCGAGAA	CTTGACTCTTCACTCTCTGG	1064
Db	958	CC	CAGTAGCTTCAAGCTGCTAGTCTCCACCGAGAA	CTTGACTCTTCACTCTCTGG	1017
Qy	1065	CCT	CTGAGCGTTACAGCTTAGGTAAATTTCAAGT	TAGTTTCTCTCTGAAACCTTCC	1124
Db	1018	CCT	CTGAGCGTTACAGCTTAGGTAAATTTCAAGT	TAGTTTCTCTCTGAAACCTTCC	1077
Qy	1125	TAC	CTGAGTACTACTTTCCCACTTAGGCTGGCTGT	GCAGAGCGCTGGAGGTGGCCAG	1184
Db	1078	TAC	CTGAGTACTACTTTCCCACTTAGGCTGGCTGT	GCAGAGCGCTGGAGGTGGCCAG	1137
Qy	1185	TT	TGTGGTTCGGAACATAAATCCCGCAGAGAGCT	GTCTATCTTCTGAAGATGGTTGCAC	1244
Db	1138	TT	TGTGGTTCGGAACATAAATCCCGCAGAGAGCT	GTCTATCTTCTGAAGATGGTTGCAC	1197
Qy	1245	CCT	CAGCGTTGGCGGAGAGATGCGCAACCGCTG	GAGAGGTGTCGCCCTCTCTCT	1304
Db	1198	CCT	CAGCGGTGTCGGGAGAGATGTCGCAACCGCT	GAGAGGTGTCGCCCTCTCTCT	1257

Qy	1305	AC	AGAGCGCTGCGGGAATCCACCACTTCGTATGAA	ACGGCTAGAAATTCCTCCAGGCT	1364
Db	1258	AC	AGAGCGCTGCGGGAATCCACCACTTCGTATGAA	ACGGCTAGAAATTCCTCCAGGCT	1317
Qy	1365	AC	AAAGGGCAAGAGTTACCCACCCCTGACCAAGAT	TAATGGCTATCATAGCCTGGAGGAG	1424
Db	1318	AC	AAAGGGCAAGAGTTACCCACCCCTGACCAAGAT	TAATGGCTATCATAGCCTGGAGGAG	1377
Qy	1425	GA	ACATAACCTTCTCCGGATGGACCAACATTTGCA	CAGATAACCCAGACAGAGGGGTG	1484
Db	1378	GA	ACATAACCTTCTCCGGATGGACCAACATTTGCA	CAGATAACCCAGACAGAGGGGTG	1437
Qy	1485	T	CCCTGCTGCACAGCGCGGAGCCCACTGAGAAA	AAACCAAAATTTGGTATTCAAGAA	1544
Db	1438	T	CCCTGCTGCACAGCGCGGAGCCCACTGAGAAA	AAACCAAAATTTGGTATTCAAGAA	1497
Qy	1545	GT	TTTCACAGAGCCCCACAGGGAACAGTCTGTTT	TGTGAATTTACCGTGGAAAAAGAAATGT	1604
Db	1498	GT	TTTCACAGAGCCCCACAGGGAACAGTCTGTTT	TGTGAATTTACCGTGGAAAAAGAAATGT	1557
Qy	1605	GA	AGAGGACCACTAATGCAACTGCACTCTCAGAT	TAGAGGAGAGCCTTCTGTTTCT	1664
Db	1558	GA	AGAGGACCACTAATGCAACTGCACTCTCAGAT	TAGAGGAGAGCCTTCTGTTTCT	1617
Qy	1665	AC	CAGACCACTTTGTAGCAAACTGATAGATTATA	TTTTGGAGGCGCCCCCAGTGAC	1724
Db	1618	AC	CAGACCACTTTGTAGCAAACTGATAGATTATA	TTTTGGAGGCGCCCCCAGTGAC	1677
Qy	1725	TT	CGAGCCAGCTCTGATTTCTGAAAGTAGGATTT	GGGGCGAGGAACCTGAGGACGATGGC	1784
Db	1678	TT	CGAGCCAGCTCTGATTTCTGAAAGTAGGATTT	GGGGCGAGGAACCTGAGGACGATGGC	1737
Qy	1785	TT	TGATAGCGATGGCTCCCTGCTCTGATCAGAC	CGTGGAAACAGGACTCGGAAGCCTTCAC	1844
Db	1738	TT	TGATAGCGATGGCTCCCTGCTCTGATCAGAC	CGTGGAAACAGGACTCGGAAGCCTTCAC	1797
Qy	1845	CT	TGGAACTCTTTCCACAGTGTAGATCTTTACA	AAACCCCAAACTTTACAGCCAGATT	1904
Db	1798	CT	TGGAACTCTTTTCCAAGTGTAGATCTTTACA	AAACCCCAAACTTTACAGCCAGATT	1857
Qy	1905	CAG	ACGCTGCCAGAAATTTGCCCCAGAGACCCAT	CAGATTCAGGAGCATCTGCTCTGGC	1964
Db	1858	CAG	ACGCTGCCAGAAATTTGCCCCAGAGACCCAT	CAGATTCAGGAGCATCTGCTCTGGC	1917
Qy	1965	AG	CTGTGTGTAGGAGCTGTCTAGGAGGAGCCCT	TTCCGGAGACCCCGACCATAGTTCC	2024
Db	1918	AG	CTGTGTGTGTAGGAGCTGTCTAGGAGGAGCC	TTCCGGAGACCCCGACCATAGTTCC	1977
Qy	2025	GG	GAGAGAGATGACTGGGAAACCGAGTGAGATGA	AGCAGAGAAATCTTTAAATTTGTGGAAC	2084
Db	1978	GG	GAGAGAGATGACTGGGAAACCGAGTGAGATGA	AGCAGAGAAATCTTTAAATTTGTGGAAC	2037
Qy	2085	T	CTTCTGTCATTCAGGAGCCCTCAACCTTTTAA	ATTTTAAAGGCTCTCTTTTCAACCG	2144
Db	2038	T	CTTCTGTCATTCAGGAGCCCTCAACCTTTTAA	ATTTTAAAGGCTCTCTTTTCAACCG	2097
Qy	2145	TC	AGGAAAGAAATTTGGAAGGCGCTCAGGACT	CTCAAGGCTCTTTCTGAGGTGAGGCT	2204
Db	2098	TC	AGGAAAGAAATTTGGAAGGCGCTCAGGACT	CTCAAGGCTCTTTCTGAGGTGAGGCT	2157
Qy	2205	T	CTCTGGGCCATCATACCTTACTTTCTTTGTA	AGCCCGCAGCTGTGTAGAGCCAGAGAT	2264
Db	2158	T	CTCTGGGCCATCATACCTTACTTTCTTTGTA	AGCCCGCAGCTGTGTAGAGCCAGAGAT	2217
Qy	2265	A	ATTGTCAGGCTGTGGGCTGGGCTGGGCTTTG	CTGGAGAAAGATACACCATATCAAG	2324
Db	2218	A	ATTGTCAGGCTGTGGGCTGGGCTGGGCTTTG	CTGGAGAAAGATACACCATATCAAG	2277
Qy	2325	AG	AAAAAAGTAACTTCTCTGGAAGAGTTACT	CTGAGTATTTATATAAGTGGTGTAGGAT	2384
Db	2278	AG	AAAAAAGTAACTTCTCTGGAAGAGTTACT	CTGAGTATTTATATAAGTGGTGTAGGAT	2337
Qy	2385	CG	AAAGGACCATTTGGGAAAGAAATTTTGAA	AGGGATGGATGACAGGTTCCAGAAACGAAATTC	2444

```

|||||
2338 CGCAAGACGACCTGGGAAGATTTTCCAAAGGATGATGCAGGTTCCAGAAACGAATTCAA 2397
|||||
QY 2445 GAAACAGAGTGGCATTTGGCTACTGCTTGGCTTTGAGCAGACAGAAAAATGTTTAAT 2504
|||||
Db 2398 GAAACAGAGTGGCATTTGGCTACTGCTTGGCTTTGAGCAGACAGAAAAATGTTTAAT 2457
|||||
QY 2505 AGACTGAGTGACGAGTCAAGAGCTTACTGTTGTACAGCAATGTTTAAAGAGTGAACAGCC 2564
|||||
Db 2458 AGACTGAGTGACGAGTCAAGAGCTTACTGTTGTACAGCAATGTTTAAAGAGTGAACAGCC 2517
|||||
QY 2565 --TGCAACCCGTCGACCTCTGTCCTTACTGAGAGTTCCTTTAAAAAACAACACTGG 2622
|||||
Db 2518 CTGCAAGCCGTCGACCTCTGTCCTTACTTGTAGAGTTTCCCTTAAAAAACAACACTGG 2577
|||||
QY 2623 CAGCTGCTCTTGGACATGTTTAAAGAAACAACACTGTATCT 2664
|||||
Db 2578 CAGCTGCTCTTGGACATGTTTAAAGAAACAACACTGTATCT 2619
|||||

RESULT 3
AK086606
LOCUS AK086606 1342 bp mRNA linear HTC 03-APR-2004
DEFINITION Mus musculus 15 days embryo head cDNA, RIKEN full-length enriched
library, clone:D930040107 product:hypothetical protein, full insert
sequence.
ACCESSION AK086606
VERSION AK086606.1 GI:26352128
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1
REFERENCE
AUTHORS Carninci,P. and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
2
REFERENCE
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159
3
REFERENCE
AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
Konno,H., Akiyama,J., Nishi,K., Kitsuai,T., Tashiro,H., Itoh,M.,
Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
Yanamoto,R., Maetumoto,H., Sakaguchi,S., Ikegami,T., Kaishwagi,K.,
Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,S., Watahiki,M.,
Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J.,
Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multipillar sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PUBMED 11076861
4
REFERENCE
AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409, 585-590 (2001)
MEDLINE 11042159
5
REFERENCE
AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
TITLE Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
JOURNAL Nature 420, 563-573 (2002)
MEDLINE 11042159
6
REFERENCE
AUTHORS Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,

```

```

Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
Hayaehida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
Hori.F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
Kato,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
Okazaki,Y., Saito,R., Saito,H., Sakai,C., Sakai,K., Sakazume,N.,
Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akaiira,S.,
Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
Muramatsu,M. and Hayashizaki,Y.
Direct Submission
Submitted (16-APR-2002) Yoshihide Hayaehizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration and Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.jp,
URL:http://genome.gsc.riken.jp/, Tel:81-45-503-9222,
Fax:81-45-503-9216)
COMMENT
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Please visit our web site for further details.
URL:http://genome.gsc.riken.jp/
URL:http://fantom.gsc.riken.jp/.
FEATURES
Location/Qualifiers
source
1. 1342
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM DB:D930040107"
/db_xref="taxon:10090"
/clone="D930040107"
/tissue_type="head"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="15 days embryo"
214. _>1341
/notes="unnamed protein product; hypothetical protein
(evidence: ProCrest)
putative"
/codon_start=1
/protein_id="BAC39701.1"
/db_xref="GI:26352129"
/tranlation="METGTHARKRPGRLGWSFRLPFLRRSHACSEFPFPSSRPNP
GNSALPERRTRYTKLLSLLQALLPSLFQKLLWSOLFGLIPTRLMDFEASYSALRA
LGRGESAAPTVOKSLSSLDSSDLIVSSLDWLEGLQWQSSDLEKLKQAEAR
LDSAAPTFLLEQQLWVELLPSSLDAGLVSRLSDSSGLSPVLSLGNPKVSVLLN
PSYLDYLPQGLRCQSSAGGQFVGFRTITPESCYLSEDCGHPQPLRAENSAWARRC
PPLSTGLPEIHLRMKRLFLQAKGQELTPDQNGYHLSLEHNLRLMDPQHCTD
NPAQVSPAADREPTKEKPELVIEVSQSPQSSSLFCLELPV"
ORIGIN
Query Match 23.2%; Score 1271; DB 3; Length 1342;
Best Local Similarity 98.5%; Pred. No. 1.1e-307;
Matches 1325; Conservative 0; Mismatches 15; Indels 5; Gaps 4;
Qy 250 GCCACCGCTTGGCGGGCCCTCTGGCGCTGCGGTGACGACACTGTTGCGGAAGCCGCGC 309
Db 2 GCCACCGCTTGGCGGGCCCTCTGGCGCTGCGGTGACGACACTGTTGCGGAAGCCGCGC 61
Qy 310 TCTCTGGGCTCTCTCTGCGCGCGGGAATCGGACTGCAGTACCCTCCGCTGCTGGGC 369
Db 62 TCTCTGGGCTCTCTCTGCGCGCGGGAATCGGACTGCAGTACCCTCCGCTGCTGGGC 121
Qy 370 AAGCGGAGACTGTGTAGACTCGGATCCAGCTGCGCTGACCGCTGAGCTGTGCTCT 429
Db 122 AAGCGGAGACTGTGTAGACTCGGATCCAGCTGCGCTGACCGCTGAGCTGTGCTCT 181
Qy 430 CCTCTCTGTGAGAACCGCCCAAGGAAGAGATGGAGACAGAACCCACAGGCCCGGA 489
Db 182 CCTCTCTGTGAGAACCGCCCAAGGAAGAGATGGAGACAGAACCCACAGGCCCGGA 241

```



Qy	490	AGCGGCTGGCCCTCGGCTGGGCTCTGGTTCCGGCTGCCCCCTTCTCTCGGGGATCGCAGC	549
Db	242	AGCGGCTGGCCCTCGGCTGGGCTCTGGTTCCGGCTGCCCCCTTCTCTCGGGGATCGCAGC	301
Qy	550	CCTGCTCTTCGGAGTTCCCGCGGCTTCTCTCGACAAATTCGCGGAACCTCCGCTCTGC	609
Db	302	CCTGCTCTTCGGAGTTCCCGCGGCTTCTCTCGACAAATTCGCGGAACCTCCGCTCTGC	361
Qy	610	CCGAGCGTCGACAGGTACTGGACCAAATTTGCTTTCTCAGCTCCTTTGCCCTGCTCCCTA	669
Db	362	CCGAGCGTCGACAGGTACTGGACCAAATTTGCTTTCTCAGCTCCTTTGCCCTGCTCCCTA	421
Qy	670	GCCTATTCCAGAACTGCTGCTTTGGAGCAGACTTTCCGGGGGCGCTGATTCCTTACCAGAT	729
Db	422	GCCTATTTCAGAACTGCTGCTTTGGAGCAGACTTTTCGGGGGCGCTGATTCCTTACCAGAT	481
Qy	730	GGCTAGATTTTTCGCGCAAGTTACAGCGCCCTGAGAGCTTTCGAGAGGACGGGAGGAATCTG	789
Db	482	GGCTAGATTTTTCGCGCAAGTTACAGCGCCCTGAGAGCTTTCGAGAGGACGGGAGGAATCTG	541
Qy	790	ACGCTCCACGGTCGAGAGTCTCTGAGTTACACTGCGGCTGGACTCTTTGCGAAGACTC	849
Db	542	CCGCTCCACGGTCGAGAGTCTCTGAGTT-CACTGCGGCTGGACTCTTTGCGAAGACTT-599	599
Qy	850	CGCTCGTCAGTACTCTTTGCATTGGCTAGAGAGGGACTCCAGTGGCAGTGCCTCCTCTCA	909
Db	600	-GGTTGTCAGTTCTCTTG-ATTGGCTAGAGAGGGACTCCAGTGGCAGTGCCTCCTCTCA	657
Qy	910	GA-CTGGAAGTTAAACTCAAGGCCACGGAAGAGCTTTAGACTCTGCAGCGCCACATTTC	968
Db	658	GACCTGAGTTTAAACTCAAGGCCACGGAAGAGCTTTAGACTCTGCAGCGCCACATTTC	717
Qy	969	CTCTGAGCAGCAGCTGTGGGAGTGGAGTTGTGCCAGTAGCCTTCAAGCTGGTCTA	1028
Db	718	CTCTGAGCAGCAGCTGTGGGAGTGGAGTTGTGCCAGTAGCCTTCAAGCTGGTCTA	777
Qy	1029	GTCTCCACCGAGAACTTGACTCTTCATCTCTGGGCTCTGAGGGTTCCAGAGCTTAGGT	1088
Db	778	GTCTCCACCGAGAACTTGACTCTTCATCTCTGGGCTCTGAGGGTTCCAGAGCTTAGGT	837
Qy	1089	AAATTCAAGGTAGTTTCTATCTCTGAAACCCCTTCTCACTGGACTACTTCTCCCAAGTTA	1148
Db	838	AAATTCAAGGTAGTTTCTATCTCTGAAACCCCTTCTCACTGGACTACTTCTCCCAAGTTA	897
Qy	1149	GGGCTGCGCTGCAGACAGCGCTGGAGGTGGCAGTTTGTGGTTTCGGAACATAACC	1208
Db	898	GGGCTGCGCTGCAGACAGCGCTGGAGGTGGCAGTTTGTGGTTTCGGAACATAACC	957
Qy	1209	CCAGAGCTGCTATCTTCTGAGATGGTTGTCAACCTCAGCCGTTCGGGCGAGAGTG	1268
Db	958	CCAGAGCTGCTATCTTCTGAGATGGTTGTCAACCTCAGCCGTTCGGGCGAGAGTG	1017
Qy	1269	TCGGCAACCGCTGGAGAGGTGTCGCTCTCTCTACAGAAAGCCCTGCCGGAATCCAC	1328
Db	1018	TCGGCAACCGCTGGAGAGGTGTCGCTCTCTCTACAGAAAGCCCTGCCGGAATCCAC	1077
Qy	1329	CACCTTCGTATGAAACCGCTAGAAATTCCTCCAGGCTAAAGGGCGAAGATTACCCACC	1388
Db	1078	CACCTTCGTATGAAACCGCTAGAAATTCCTCCAGGCTAAAGGGCGAAGATTACCCACC	1137
Qy	1389	CCTGACCAAGATAATGCTATCATAGCTGGAGAGGAAATAATACCTTCTCCGGATGGAC	1448
Db	1138	CCTGACCAAGATAATGCTATCATAGCTGGAGAGGAAATAATACCTTCTCCGGATGGAC	1197
Qy	1449	CCACAACTTGCAAGATAAACCAGCAGACAGGCGGTGTCCCTGCTGCAGACAGGCGGAG	1508
Db	1198	CCACAACTTGCAAGATAAACCAGCAGACAGGCGGTGTCCCTGCTGCAGACAGGCGGAG	1257
Qy	1509	CCCACTGAGAAAAACCGAATTTGGTGAATTCAGAGTTTTCACAGGCCCCAGGGAGC	1568
Db	1258	CCCACTGAGAAAAACCGAATTTGGTGAATTCAGAGTTTTCACAGGCCCCAGGGAGC	1317

QY	1569	AGTCTGTTTGTGAATTACCGTGG	1593	
DB	1318	AGTCTGTTTGTGAATTACCGTGG	1342	
RESULT 4				
BQ922494				
LOCUS				
DEFINITION	B0922494	871 bp mRNA linear EST 20-AUG-2002		
	AGENCOURT 8921919 NCI_CGAP Co24 Mus musculus cDNA clone			
	IMAGE:6395811 5', mRNA sequence.			
ACCESSION	B0922494			
VERSION	BQ922494.1	GI:22337525		
KEYWORDS	EST.			
SOURCE	Mus musculus (house mouse)			
ORGANISM	Mus musculus			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
TITLE	1 (bases 1 to 871)			
JOURNAL	NIH-MGC <a href="http://mgc.nci.nih.gov/">http://mgc.nci.nih.gov/</a> .			
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: <a href="mailto:cgapbs-t@mail.nih.gov">cgapbs-t@mail.nih.gov</a> Tissue Procurement: The Cepko Laboratory cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <a href="http://image.llnl.gov">http://image.llnl.gov</a> Plate: LLAM13891 row: b column: 04 High quality sequence stop: 594.			
FEATURES				
source	1..871			
	/organism="Mus musculus"			
	/mol_type="mRNA"			
	/strain="FVB/N"			
	/db_xref="taxon:10090"			
	/clone="IMAGE:6395811"			
	/lab_host="DH10B (Ti phage-resistant)"			
	/clone_lib="NCI_CGAP_Co24"			
	/note="Organ: colon; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.6 kb. Constructed by Life Technologies. Note: this is a NCI_CGAP Library."			
ORIGIN				
Query Match	15.1%;	Score 828.2;	DB 5;	Length 871;
Best Local Similarity	97.9%;	Pred. No. 2.2e-196;		
Matches 839;	Conservative 0;	Mismatches 18;	Indels 0;	Gaps 0;
QY	4463	TTCAAGTTAGCGGTACACCTGTGTGCATGGTTGAGTAAATGTTACCAAGTTCTTTAGTGGT	4522	
DB	1	TTGTGTTTGTGTACAGCTGTGTGTGCATGGTTGAGTAAATGTTACCAAGTTCTTTAGTGGT	60	
QY	4523	TACATGCACAAAAGAGAGAGACCTCTGAGTCGGGTGTGGGATGAGCTTCCAGACCTGGCA	4582	
DB	61	TACATGCACAAAAGAGAGAGACCTCTGAGTCGGGTGTGGGATGAGCTTCCAGACCTGGCA	120	
QY	4583	GGGTAAACTACCTCAGTTTATAATCTCCCTGGTTATTTCCGTTTGATGTGATCTAAGGTC	4642	
DB	121	GGGTAAACTACCTCAGTTTATAATCTCCCTGGTTATTTCCGTTTGATGTGATCTAAGGTC	180	
QY	4643	TGCCTCAGTCGTGATGATGTTTCCACACACAAGGTTAGTAAAGTGCACGACCCAGAAA	4702	
DB	181	TGCCTCAGTCGTGATGATGTTTCCACACACAAGGTTAGTAAAGTGCACGACCCAGAAA	240	
QY	4703	CGTTGGTCTTATTTTGGAGAACCCCAATTTCTGTGATTTTATGCACCTGCCCTTAGTGA	4762	
DB	241	CGTTGGTCTTATTTTGGAGAACCCCAATTTCTGTGATTTTATGCACCTGCCCTTAGTGA	300	
QY	4763	ACTCCAGATGCATTAAAGAGTCTGGTTAGTCCCGTGGGAATGGGCTAGTTTAGAAGCT	4822	



Db	301	ACTCCAGATGCAATTAAGAGCTCTGGTTAGTCCGTGGGAATGGCTAGTTTGAAGCT	360
Qy	4823	ATGTTTGGAAAGCAGCGAAGTTGACTTTAGGAAGAAAAGCTGTGACAGTGTAGACATT	4882
Db	361	ATGTTTGGAAAGCAGCGAAGTTGACTTTAGGAAGAAAAGCTGTGACAGTGTAGACATT	420
Qy	4883	TCTTTTAAACCGGACTGCGAGCTTAAACAACACTTGATTTTTCAGATGATTAGGTTTGTTC	4942
Db	421	TCTTTTAAACCGGACTGCGAGCTTAAACAACACTTGATTTTTCAGATGATTAGGTTTGTTC	480
Qy	4943	TCAGACCCAGCAGCCTGTATATTTTAAAAATTTGTCAGATTACACCTTCACATCAAAATGA	5002
Db	481	TCAGACCCAGCAGCCTGTATATTTTAAAAATTTGTCAGATTACACCTTCACATCAAAATGA	540
Qy	5003	GTAAATGACTCATGCTCGAGACATGCTCTGATGTGGCAAGAACAGAAAGATCTTTGAC	5062
Db	541	GTAAATGACTCATGCTCGAGACATGCTCTGATGTGGCAAGAACAGAAAGATCTTTGAC	600
Qy	5063	TGAAGGAGAAAACCTGTCTATTCATCCAGCCCCCAGGAAAGAACACCTTCAAGGCAGG	5122
Db	601	TGAAGGAGAAAACCTGTCTATTCATCCAGCCCCCAGGAAAGAACACCTTCAAGGCAGG	660
Qy	5123	CAGGACGAGCAGCAGCATGTGTTCTAGTTGAATACATTCAGTCTTGCAGTGGTG	5182
Db	661	CAGGACGAGCAGCAGCATGTGTTCTAGTTGAATACATTCAGTCTTGCAGTGGTG	720
Qy	5183	CTTTAGATCTGTGTAGCATGTGAGGCTCTGTACAGTGGGGCCACACTTCTGAGGGCTG	5242
Db	721	CTTTAGATCTGTGTAGCATGTGAGGCTCTGTACAGTGGGGCCACACTTCTGAGGGCTG	780
Qy	5243	AAATGTGCAACCCCTTTATCTAACTTGAAATCAAAACCGTCAAAATTTTATATAT	5302
Db	781	AAATGTGCAACCCCTTTATCTAACTTGAAATCAAAACCGTCAAAATTTTATATAT	840
Qy	5303	TTAGAAAGAGTGGGG	5319
Db	841	TTTAGGAAAAGAGTGG	857
RESULT 5			
LOCUS	BQ926670	966 bp	mRNA
DEFINITION	AGENCOURT_8815092 NCI_CGAP Co24	Mus musculus	linear
ACCESSION	BQ926670	IMAGE:6398165	5', mRNA sequence.
VERSION	BQ926670.1	GI:22341701	
KEYWORDS	EST.		
SOURCE	Mus musculus		
ORGANISM	Mus musculus (house mouse)		
REFERENCE	1	(bases 1 to 966)	
AUTHORS	NIH-MGC	http://mgc.nci.nih.gov/.	
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)		
JOURNAL	Unpublished (1999)		
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgapbs-remail.nih.gov cDNA Library Preparation: The Cepko Laboratory cDNA Sequencing Arrayed by: The I.M.A.G.E. Consortium (ILNL) Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/ILNL at: http://image.llnl.gov Plate: LLNL13897 High quality sequence scop: 613. Location/Qualifiers		
FEATURES	1..966		
source	/organism="Mus musculus"		
	/mol_type="mRNA"		
	/strain="FVB/N"		
	/db_xref="taxon:10090"		
	/clone="IMAGE:6398165"		

/lab_host="DH10B (T1 phage-resistant)"			
/clone_lib="NCI_CGAP Co24"			
/note="Organ: colon; Vector: pCMV-SPORT6; Site 1: NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.6 kb. Constructed by Life Technologies. Note: this is a NCI_CGAP Library."			
ORIGIN			
Query Match	14.8%	Score 810;	DB 5; Length 966;
Best Local Similarity	96.4%	Pred. No. 8.3e-192;	
Matches	873;	Conservative	0; Mismatches 25; Indels 8; Gaps 4;
Qy	3935	GAAAGCGGGTGGTGTCTTTTCAAAAGGCACACACATTATGCGCCAGTGGTCTGCGAA	3994
Db	1	GAAAGCGGGTGGTGTCTTTTCAAAAGGCACACACATTATGCGCCAGTGGTCTGCGAA	59
Qy	3995	ACTTTAAACATATTTGGTTTCTGTGCAAACTGTCCAAATTTGGAGGTTTCTGTGTGTGTAGT	4054
Db	60	ACTTTAAACATATTTGGTTTCTGTGCAAACTGTCCAAATTTGGAGGTTTCTGTGTGTGTAGT	119
Qy	4055	TTCAAGTTGAATGTGGCTCTCTGTAGCTTTCAGAAATGGGTAGAAATCATAAAGCACTCTT	4114
Db	120	TTCAAGTTGAATGTGGCTCTCTGTAGCTTTCAGAAATGGGTAGAAATCATAAAGCACTCTT	179
Qy	4115	AAGTAATCATGTGCATGTAGACATTTTTTTTTTAACTAGGGGTATTTGGGAGACATG	4174
Db	180	AAGTAATCATGTGCATGTAGACATTTTTTTTTTAACTAGGGGTATTTGGGAGACATG	237
Qy	4175	TCAGATATTCACATTTTGTATTTATGTGCTTTAAACACAGTGTACTTACACCTGCCCTCAG	4234
Db	238	TCAGATATTCACATTTTGTATTTATGTGCTTTAAACACAGTGTACTTACACCTGCCCTCAG	297
Qy	4235	CACGTGGCACTTTTCAAAACCTGTCTTGGAGACACCTGTAAACCTTGGATGGTCAAGTCTCGGG	4294
Db	298	CACGTGGCACTTTTCAAAACCTGTCTTGGAGACACCTGTAAACCTTGGATGGTCAAGTCTCGGG	357
Qy	4295	TTTTTCATGTGTAAATGTCAACTACAAAGGTCAATATCCAGGTTTGTGTGTCTACCTT	4354
Db	358	TTTTTCATGTGTAAATGTCAACTACAAAGGTCAATATCCAGGTTTGTGTGTCTACCTT	417
Qy	4355	ATGTAGAGAGACAAAGCAGAAAGGCAGATAGAGCAGCCAGAAAGTGTAGTGTCCCC	4414
Db	418	ATGTAGAGAGAGACAAAGCAGAAAGGCAGATAGAGCAGCCAGAAAGTGTAGTGTCCCC	477
Qy	4415	CCCAAGCGTCTCTTAGATAGTGTGTGGATCAGGTGTCTTGGTTTGTTCAGTTAGCGT	4474
Db	478	CCCAAGCGTCTCTTAGATAGTGTGTGGATCAGGTGTCTTGGTTTGTTCAGTTAGCGT	537
Qy	4475	CACCTGTGTGCATGGTTGAGGTAATGTGTTCACAGTTCTTTAGTGGTTACATGCACAA	4534
Db	538	CACCTGTGTGCATGGTTGAGGTAATGTGTTCACAGTTCTTTAGTGGTTACATGCACAA	597
Qy	4535	AGAGAGACCTCTGAGTCGGGTGTGGGATGAGCTTTCCAGACCTGGCAGGGAATACCTACC	4594
Db	598	AGAGAGACCTCTGAGTCGGGTGTGGGATGAGCTTTCCAGACCTGGCAGGGAATACCTACC	657
Qy	4595	TCAGTTTATAATCTCCCTGGTTATTTCCGTTTGAATGTGATCTAAGGTCGTGCTCAGTGGT	4654
Db	658	TCAGTTTATAATCTCCCTGGTTATTTCCGTTTGAATGTGATCTAAGGTCGTGCTCAGTGGT	717
Qy	4655	GATGATGTCATCACAACAAGGTTAGTAAGTGTGACGACGACGAAACGTTGTCTTAT	4714
Db	718	GATGATGTCATCACAACAAGGTTAGTAAGTGTGACGACGACGAAACGTTGTGTCTTAT	777
Qy	4715	TTTTTGAGAACCCCATTTCTGTGTATTTTATGCA-CCTGGCTTTTGTAGTGAATCCAGAGTG	4773
Db	778	TTTTTGAGAACCCCATTTCTGTGTATTTTATGCAACCTTGTCTTGTAGTGAATCCAGAGTG	837
Qy	4774	CATTAAAGAGTCTGG-----TTTAGTGCCTGGGAATGGGCTAGTTTGAAGCTATGTTTG	4829
Db	838	CATTAAAGAGTCTGGGTTTAAATGGCTGGGGAATGGGCTACTTTACAAACCTATTTTTT	897
Qy	4830	GAAAGC	4835

```
Db      898  GGAAC 903

RESULT 6
AV055059
LOCUS   AV055059 997 bp mRNA linear EST 24-OCT-2001
DEFINITION AV055059 Mus musculus pancreas C57BL/6J adult Mus musculus cDNA
clone 1810033K10, mRNA sequence.
ACCESSION AV055059
VERSION   AV055059.2 GI:16380666
KEYWORDS  EST.
SOURCE   Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE
AUTHORS  Arakawa, T., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T.,
Hara, A., Hiramoto, K., Hori, P., Ishii, Y., Ito, M., Kawai, J.,
Konno, H., Kouda, M., Koya, S., Matsuyama, T., Miyazaki, A., Nomura, K.,
Ohno, M., Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K.,
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, P.,
Takeda, Y., Tanaka, T., Toya, T., Muramatsu, M. and Hayaishizaki, Y.
RIKEN Mouse ESTs (Arakawa, T., et al. 2001)
Unpublished (2001)
On Jun 23, 1999 this sequence version replaced gi:5154806.
Contact: Yoshihide Hayaishizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayaishizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi, K., Fujiwaki, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,
Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T.,
Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A.
and Hayaishizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P.,
Sugahara, Y. and Hayaishizaki, Y.
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Kondo, S., Shinagawa, A., Saito, T., Kiyosawa, H., Yamanaka, I.,
Aizawa, K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K. and
Hayaishizaki, Y.
Computational Analysis of Full-length Mouse cDNAs Compared with
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp/) for
further details.
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
FEATURES             Location/Qualifiers
     source           1..997
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="1810033K10"
                     /sex="male"
                     /tissue_type="pancreas"
```

```

/dev_stage="adult"
/clone_lib="Mus musculus pancreas C57BL/6J adult"

ORIGIN
Query Match      14.8%   Score 808.6;   DB 1;   Length 997;
Best Local Similarity 96.4%;   Pred. No. 1.9e-191;
Matches 834;   Conservative 0;   Mismatches 30;   Indels 1;   Gaps 1;

Qy 4584 GGTAAACTACCTCAGTTTAAATCTCCCTGGTTATTTCCTGTTGATGATCTAAGGTCT 4643
Db      |||
Qy 4644 GCCTCAGTGTGATGATGTTTCATCCACACACAGAGTTAGTAAGAGTGCACGACGAGAAC 4703
Db      |||
Qy 4704 GTTGGTCTTATTTTTGAGAACCCCAATTTCTGTGTATTTTATGCCTGCTTTTAGTGAA 4763
Db      |||
Qy 4764 CTCAGAGTGCATTAAAGAGTCTGGTTTGTGCGGTGGGAATGGGCTAGTTTAAAGACTA 4823
Db      |||
Qy 4824 TGTGGAAGAGCAGGCAAGTTGACTTTAGGAAGAAAGCTGTGCACAGTGTGTAGACATTT 4883
Db      |||
Qy 4884 CTTTAAACCGGAGTGCAGCTTAAACACACTTCAATTTTTCAGATGATAGGTTTGTGTTCT 4943
Db      |||
Qy 4943 GAGACCCAGCACCTGTATATTTTAAATTTTGTTCAGATTTACACCTTCAATGAG 5003
Db      |||
Qy 5004 TAAATGACTCATGCTGCAGACATGCTCTGATGGTGGCAAGAACAGAGATCTTTGACT 5063
Db      |||
Qy 5064 GAAGGAGAAAACGTGTCATTTTCCAGATTTACACCTTCAACCTTCAAGGCAAGGC 5123
Db      |||
Qy 5124 AGCAGCAGCAGGAGGATGTTGTTCTTAGTTGAATACACATTTCAAGTCTTGCAGTGTGC 5183
Db      |||
Qy 5184 TTTAGATCTGTGATGATGAGGCTCTGTACAGTGGGGGCCACACTTCTGAGGGCTGA 5243
Db      |||
Qy 5244 AATGTGCAACCTTTTATCTAACTTGAATCAAAACCGTCAAAATTTTATTTTATAATT 5303
Db      |||
Qy 5304 TAAAGAAGAGTGGGGAATGACATTTTTCAGTTGGCTTTTTCAGCTCAGTCAATTTTACG 5363
Db      |||
Qy 5364 TGTAACTGTGAGATTTGATAGTCAAGATATATTTGTATATATATTAATCTAATCTGTA 5423
Db      |||
Qy 5424 AATTGTAATAATATATTTTGCATTT 5448
Db      |||
Qy 5448 AATTGTAATAATATATTTTGCATTT 5497
Db      |||

RESULT 7
CF583619/c
LOCUS   CF583619 890 bp mRNA linear EST 24-SEP-2003
DEFINITION AGENCOURT_8845828_updated NIH_MGC_137 Mus musculus cDNA clone
```

IMAGE:6431877 3', mRNA sequence.					
ACCESSION	CF583619				
VERSION	CF583619.1	GI:35196881			
KEYWORDS	EST.				
SOURCE	Mus musculus (house mouse)				
ORGANISM	Mus musculus				
REFERENCE	NIH-MGC <a href="http://mgs.nci.nih.gov/">http://mgs.nci.nih.gov/</a>				
AUTHORS	National Institutes of Health, Mammalian Gene Collection (MGC)				
TITLE	Unpublished (1999)				
JOURNAL	Contact: Daniela S. Gerhard, Ph.D.				
COMMENT	Office of Cancer Genomics National Cancer Institute / NIH Bldg. 31 Rm10A07 Bethesda, MD 20892 Email: <a href="mailto:cgapbs-r@mail.nih.gov">cgapbs-r@mail.nih.gov</a> Tissue Procurement: Gerard Gradwohl (PNAS 97 P1607-1611, 2000) cDNA Library Preparation: Catherine Lee, Endocrine Pancreas Consortium cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <a href="http://image.llnl.gov">http://image.llnl.gov</a> Plate: IRBD20 row: h column: 10 High quality sequence stop: 675.				
FEATURES	Location/Qualifiers				
source	1..890				
	/organism="Mus musculus"				
	/mol_type="mRNA"				
	/db_xref="taxon:10090"				
	/clone="IMAGE:6431877"				
	/lab_host="DH10B"				
	/clone_lib="NIH_MGC_137"				
	/note="Organ: pancreas; Vector: pSPORT1; Site1: SalI; Site2: NotI; Library consists of a pool of clones rearrayed from the following libraries: Melton normalized mixed mouse pancreas 1 NI-MMS1, Amplified Melton mouse islets 1 MSL-A, and Kaestner ngm3 wt. Clones rearrayed in the laboratory of K. Kaestner (University of Pennsylvania). Note: this is a NIH_MGC Library."				
ORIGIN					
Query Match	14.6%;	Score 799.2;	DB 7; Length 890;		
Best Local Similarity	99.0%;	Pred. No. 4.2e-189;			
Matches	857;	Conservative 0;	Mismatches 3; Indels 6; Gaps 5;		
QY	4609	CCCTGGTTATTCGGTTTGATGTGATCTPAA-GGTCTGCCTCAGT-GGTGATGATGTTTCAT	4666		
DB	867	CCCTGGTTATTCGGTTTGATGTGATCTPAAAGGTCCTGCCTCAGTGGGTGATGATGTTTCAT	808		
QY	4667	CCACACACAAGG--TTAGTAGAGTGCACGACCCAGAACGTT-GGTCTTATTTTGAGAA	4723		
DB	807	CCACACACAAGGTTTGTAGTAGTGCACGACCCAGAACGTTGGGTCTTATTTTGAGAA	748		
QY	4724	CCCCCATTTCT-GTGTATTTTATGCACTGCTCTTAGTGAACCTCCAGAGTGCATTAAAGA	4782		
DB	747	CCCCCATTTCTGTTATTTTATGCACTGCTCTTAGTGAACCTCCAGAGTGCATTAAAGA	688		
QY	4783	GTCTGGTTTATGTCGGTGGGAATGGCTAGTTTGAAGCTATGTTTGAAGAGCCAGGCAAG	4842		
DB	687	GTCTGGTTTATGTCGGTGGGAATGGCTAGTTTGAAGCTATGTTTGAAGAGCCAGGCAAG	628		
QY	4843	TTGACTTTTAGGAAGAAAAGCTGTGACAGTGTGACAGATTTCTTTTAAACCGGACTGCAG	4902		
DB	627	TTGACTTTTAGGAAGAAAAGCTGTGACAGTGTGACAGATTTCTTTTAAACCGGACTGCAG	568		
QY	4903	CTTAACAACACTGTATTTTCAGATGATTTAGGTTTTTTGTTCTGAGACCCAGCAGCTGTATA	4962		
DB	567	CTTAACAACACTGTATTTTCAGATGATTTAGGTTTTTTGTTCTGAGACCCAGCAGCTGTATA	508		
QY	4963	TTTAAAAAATGTTCCAGATTACCTTCATCTATCAAAATGAGTAAATGACTCATCGCTGCA	5022		
Db	507	TTTAAAAAATGTTCCAGATTACCTTTCACTATCAAAATGAGTAAATGACTCATCGCTGCA	448		
QY	5023	GACATGTCCTGATGGTGGCAGAACACAGAGGATCTTTGACTGAAGGAGAAAAACTGTGCAT	5082		
Db	447	GACATGTCCTGATGGTGGCAGAACACAGAGGATCTTTGACTGAAGGAGAAAAACTGTGCAT	388		
QY	5083	TGTCATCCAGCCGCCAGGAAAGAACACCTCCAGGCGAGGCGAGGCGAGGCGAGGCGATG	5142		
Db	387	TGTCATCCAGCCGCCAGGAAAGAACACCTCCAGGCGAGGCGAGGCGAGGCGAGGCGATG	328		
QY	5143	GTGGTTCTAGTTGAATACACATTCACATCTTGCAGTGGTGTCTTTAGATCTGTGTAGCATG	5202		
Db	327	GTGGTTCTAGTTGAATACACATTCACATCTTGCAGTGGTGTCTTTAGATCTGTGTAGCATG	268		
QY	5203	TGAGGCTCTGTACAGGTGGGGGCCACACCTCTGAGGGCTGAAATGTGCGAACCCCTTTATC	5262		
Db	267	TGAGGCTCTGTACAGGTGGGGGCCACACCTCTGAGGGCTGAAATGTGCGAACCCCTTTATC	208		
QY	5263	TAACTTGAATCAAAACCGTCAAAATTTTATTTTATTAATTTTAAAGAAAGAGTTGGGGAAT	5322		
Db	207	TAACTTGAATCAAAACCGTCAAAATTTTATTTTATTAATTTTAAAGAAAGAGTTGGGGAAT	148		
QY	5323	GACATTTTTCAGTTGGCTTTTCAGCTCAGTCATTTTACGTGTAAAGCTGTGAGATTTGAT	5382		
Db	147	GACATTTTTCAGTTGGCTTTTCAGCTCAGTCATTTTACGTGTAAAGCTGTGAGATTTGAT	88		
QY	5383	AGCTCAGATTATTTTGTATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTT	5442		
Db	87	AGCTCAGATTATTTTGTATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTT	28		
QY	5443	GCAATTTTAAAAAAGGAAAAAAGGAAAAA 5468			
Db	27	GCAATTTTAAAAAAGGAAAAAAGGAAAAA 2			
RESULT 8					
BB610365					
LOCUS					
DEFINITION	BB610365 RIKEN full-length enriched, 10 day old male pancreas Mus				
ACCSSION	BB610365				
VERSION	BB610365.1				
KEYWORDS	EST.				
SOURCE	Mus musculus (house mouse)				
ORGANISM					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
COMMENT					

Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.  
 RIKEN integrated sequence analysis (RISA) system--384-format  
 sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11) 1757-1771 (2000)  
 Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara, Y., and Hayashizaki, Y.  
 Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
 Kondo, S., Shinagawa, A., Saito, T., Kiyosawa, H., Yamanaka, I., Aizawa, K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K., and Hayashizaki, Y.  
 Computational Analysis of Full-length Mouse cDNAs Compared with Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)  
 Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.  
 e mouse tissues.

## FEATURES

## source

Location/Qualifiers  
 1. .1011  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="1810033K10"  
 /sex="male"  
 /tissue\_type="pancreas"  
 /dev\_stage="10 day old"  
 /clone\_lib="RIKEN full-length enriched, 10 day old male pancreas"

## ORIGIN

Query Match 14.6%; Score 796; DB 2; Length 1011;  
 Best Local Similarity 90.9%; Pred. No. 2.8e-188;  
 Matches 924; Conservative 0; Mismatches 83; Indels 10; Gaps 7;  
 3724 TTTCTTTTTTAAAGACATGGCTTATATAGTCTAGGCTGGCCTTAACTCACTACTTA 3783  
 Db 1 TTTCTTTTTTAAAGACATGGCTTATATAGTCTAGGCTGGCCTTAACTCACTACTTA 60  
 3784 GCCAAGATGGCTTTGAAGATCTCTGTCCTCTTCTTGGCTTGCACATGGCATGTC 3843  
 Db 61 GCCAAGATGGCTTTGAAGATCTCTGTCCTCTTCTTGGCTTGCACATGGCATGTC 120  
 3844 CATACATCCAGATTTTCTCTGTATCTAGTCTTTTAAAGATATTTCAAGGACTGCCCA 3903  
 Db 121 CATACATCCAGATTTTCTCTGTATCTAGTCTTTTAAAGATATTTCAAGGACTGCCCA 180  
 3904 CAATCCACAGTGCAGTACTTTCTGTACTGGGAAGCGGGTGGTGTCTTTTCAAGG 3963  
 Db 181 CAATCCACAGTGCAGTACTTTCTGTACTGGGAAGCGGGTGGTGTCTTTTCAAGG 240  
 3964 CACACACATATGCCAGTGGTCTGCCAACTTAACATATTTGGTTCTGTGCAACT 4023  
 Db 241 CACACACATATGCCAGTGGTCTGCCAACTTAACATATTTGGTTCTGTGCAACT 300  
 4024 GTCCATTTGGAAGGTTTCTGTGTCTGTAGTTTTCAGTTGAATGGTCTCTGTAGCTTT 4083  
 Db 301 GTCCATTTGGAAGGTTTCTGTGTCTGTAGTTTTCAGTTGAATGGTCTCTGTAGCTTT 360  
 4084 CAAGAATGGGTAGAAATCATAAAGCACTCTTAACTAATCAATTTGATGTAGACATTTTTT 4143  
 Db 361 CAAGAATGGGTAGAAATCATAAAGCACTCTTAACTAATCAATTTGATGTAGACATTTTTT 419  
 4144 TTTTAACTAGGGGGTATTTGGGAGACATGTACATATTCACATTTTGTATGTCTCT 4203  
 Db 420 TTTTAACTAGGGGGTATTTGGGAGACATGTACATATTCACATTTTGTATGTCTCT 479  
 4204 TAAACACAGTCTTACTTACACCTCGCTCAGCACTGGCACATTTTCAACCTGTCTTGAG 4263  
 Db 480 TAAACACAGTCTTACTTACACCTCGCTCAGCACTGGCACATTTTCAACCTGTCTTGAG 539  
 4264 ACACGTAAACTTGGATGGTGCAGTCTTCTGGGTTTTTCATGTGTAAATGTCAACTACAAAG 4323

Db 540 ACACGTAAACTTGGATGGTGCAGTTCTGGGTTTTTCATGTGTAAATGTTAACACAAAG 599  
 Qy 4324 GTCAATATCCAGGTTTGTGTGTCTTACCTTATGTAGAGAGAAAGCAAGCAAGAGGCA 4383  
 Db 600 GTCATATATCCAGGTTTGTGTGTCTTACCTTATGTAGAGAGAAAGCAAGCAAGAGGCA 659  
 Qy 4384 GATAGAGAGCCCAAGAGTGTAGT--GTCCCCCCCCAAAGCGTCTTCTAGATAGTGTGT 4441  
 Db 660 GATAGAGAGCCGACAGATAGTGTAGTTGTTCCTCCGCCCAAGACGCTTATAGATAGTGTGT 719  
 Qy 4442 CGATCAGGT 4501  
 Db 720 GGATCAGGT 779  
 Qy 4502 TGTTCACAGTCTTTTGT 4561  
 Db 780 TGTGTACAGTCTTTTGT 836  
 Qy 4562 ATGAGCTTTTCCAGACCTGGCAGGGTAACTTACCTCAGTTTATATATCTCCCTGTGTATTTTC 4621  
 Db 837 GATGAGCTTCCAGACACACGCGGAGGTAGACTATCTCAGTTAAT-ATCTCCCTGTATATT 895  
 Qy 4622 CGTTTGTATGTATCTAAGGTCTGCTCAGTGGTGTATGTATCTCAGTCCACACACAGGTTA 4681  
 Db 896 CGCTTGAAGCGATCTAATG-CTGCCTCAGTGTGTGAAGACGTTTCTAT-CACACACAGGCTA 953  
 Qy 4682 GTAAGAGTGCACACACAGAAACGTTGTGTCTTATTTTGTAGAACCCCATTTCTGTGT 4738  
 Db 954 GNAATAGTGCACGA-CAGAGAGTTGGCTATTTTATAGAACCCCTCTCTGTGTGT 1009

## RESULT 9

## CF583620

## LOCUS

## DEFINITION

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

CF583620 913 bp mRNA linear EST 24-SEP-2003  
 AGENCOURT\_11362879 updated NIH\_MGC\_137 Mus musculus CDNA clone  
 IMAGE:6431877 5', mRNA sequence.

CF583620  
 CF583620.1 GI:35196882

EST.  
 Mus musculus (house mouse)

Mus musculus  
 Rukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 913)  
 NIH-MGC <http://mgi.nci.nih.gov/>

National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)

Contact: Daniela S. Gerhard, Ph.D.  
 Office of Cancer Genomics  
 National Cancer Institute / NIH

Bldg. 31 Rm10A07 Bethesda, MD 20892  
 Email: [cgabs@mail.nih.gov](mailto:cgabs@mail.nih.gov)

Tissue procurement: Gerard Gradwohl (PNAS 97 P1607-1611, 2000)  
 cDNA Library Preparation: Catherine Lee, Endocrine Pancreas Consortium

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>  
 Plate: IRBD20 row: h column: 10

High quality sequence stop: 672.  
 Location/Qualifiers

1. .913  
 /organism="Mus musculus"  
 /mol\_type="mRNA"

/db\_xref="taxon:10090"  
 /clone="IMAGE:6431877"

/lab\_host="DH10B"  
 /clone\_lib="NIH\_MGC\_137"

/notes="Organ: pancreas; Vector: pSPORT1; Site 1: SalI;  
 Site 2: NotI; Library consists of a pool of clones"

## FEATURES

## source

rearrayed from the following libraries: Melton normalized mixed mouse pancreas 1 NI-WMS1, Amplified Melton mouse islets 1 MSL-A, and Kaetner ngn3 wt. Clones rearrayed in the laboratory of K. Kaetner (University of Pennsylvania). Note: this is a NIH\_MGC Library."

ORIGIN		Query Match		14.5%; Score 794.2; DB 7; Length 913;		Best Local Similarity		97.8%; Pred. No. 7.7e-188;		Matches 805; Conservative 0; Mismatches 18; Indels 0; Gaps 0;	
QY	4304	GTAAATGTCAACTACAAAGTCAATATCCAGTTTGTGTCTTACCTTATGTAGAGA	4363			Db	11	GTAAATGTCAACTACAAAGTCAATATCCAGTTTGTGTCTTACCTTATGTAGAGA	70		
QY	4364	GAAGCAAGCAGAAAGGCGAGTAGACGAGCCGAGAGTGTAGTGTCCCGCCCAAGCG	4423			Db	71	GAAGCAAGCAGAAAGGCGAGTAGACGAGCCGAGAGTGTAGTGTCCCGCCCAAGCG	130		
QY	4424	TGCTTCTAGATAGTGTGGATCAGGTGTCTTGTGTGTTCAGTTAGCTCAGCTGTG	4483			Db	131	TGCTTCTAGATAGTGTGGATCAGGTGTCTTGTGTGTTCAGTTAGCTCAGCTGTG	190		
QY	4484	TGATGTGTGAGTTAATGTGTACAGTTCCTTTAGTGTGTATCATGACAAAGAGAGAC	4543			Db	191	TGATGTGTGAGTTAATGTGTACAGTTCCTTTAGTGTGTATCATGACAAAGAGAGAC	250		
QY	4544	CTCTGAGTGGGTGGGATGAGCTTCCAGACCTGGCAGGTAACCTCAGTTTAT	4603			Db	251	CTCTGAGTGGGTGGGATGAGCTTCCAGACCTGGCAGGTAACCTCAGTTTAT	310		
QY	4604	AATCTCCCTGGTTATTCCTGTTGATGTATAGTGTCTGCTCAGTGGTGTATGATGT	4663			Db	311	AATCTCCCTGGTTATTCCTGTTGATGTATAGTGTCTGCTCAGTGGTGTATGATGT	370		
QY	4664	CATCACACACAAGTTAGTAAAGTGCACGACGAGAAACGTTGCTTATTTTGTAGAA	4723			Db	371	CATCACACACAAGTTAGTAAAGTGCACGACGAGAAACGTTGCTTATTTTGTAGAA	430		
QY	4724	CCCCCATTTCTGTGTATTTATGACCTGCCTTTAGTGAATCCAGAGTGCATTAAGAG	4783			Db	431	CCCCCATTTCTGTGTATTTATGACCTGCCTTTAGTGAATCCAGAGTGCATTAAGAG	490		
QY	4784	TCTGTTTGTAGTCCGCTGGGATGGCTAGTTTGTAGAGCTATGTTTGGAAAGCAGCAGT	4843			Db	491	TCTGTTTGTAGTCCGCTGGGATGGCTAGTTTGTAGAGCTATGTTTGGAAAGCAGCAGT	550		
QY	4844	TGACTTTAGGAAGAAAGCTGTGACAGTGTGTAGACATTTCTTTTAAACCGGACTGCAGC	4903			Db	551	TGACTTTAGGAAGAAAGCTGTGACAGTGTGTAGACATTTCTTTTAAACCGGACTGCAGC	610		
QY	4904	TTAAACAACATTTTTCAGATGATTTAGGTTTGTGTTCTGAGACCCAGCAGCTGTATAT	4963			Db	611	TTAAACAACATTTTTCAGATGATTTAGGTTTGTGTTCTGAGACCCAGCAGCTGTATAT	670		
QY	4964	TTAAAAATTGTTCCAGATTACACCTTCACTATCAAAATGAGTAAATGACTCATGCTGCAG	5023			Db	671	TTAAAAATTGTTCCAGATTACACCTTCACTATCAAAATGAGTAAATGACTCATGCTGCAG	730		
QY	5024	ACATGCTCTGATGGTGGCAAGAACAGAGATCTTTTGACTGAAGAGAGAAAACTGTCAAT	5083			Db	731	ACATGCTCTGATGGTGGCAAGAACAGAGATCTTTTGACTGAAGAGAGAAAACTGTCAAT	790		
QY	5084	GTCATCCAGCCCCCAGGAGAGACACCTCCAGGCGAGG	5126			Db	791	GTCATCCAGCCCCCAGGAGAGACACCTCCAGGCGAGG	833		

RESULT 10  
CA978373  
LOCUS CA978373 946 bp mRNA linear EST 06-JAN-2003  
DEFINITION AGNCOURT\_11277002 NIH\_MGC\_164 Mus musculus cDNA clone  
IMAGE:30144607 5', mRNA sequence.

CA978373  
EST. CA978373.1 GI:27511027  
Mus musculus (house mouse)  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
NIH-MGC http://mgi.nci.nih.gov/  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgaaps-r@mail.nih.gov  
Tissue Procurement: Dr. David Rowe and Dr. Mina  
cDNA Library Preparation: Invitrogen Corp  
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: NDAM0057 row: f column: 08  
High quality sequence start: 24  
High quality sequence stop: 693.  
Location/Qualifiers  
1. 946  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="IMAGE:30144607"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH\_MGC\_164"  
/note="Vector: pCMV-SPORT6.1; Site 1: EcoRV; Site 2: NotI; Non-normalized full-length enriched library from pooled mouse embryonic limb, maxilla and mandible, day 10.5 and 11.5 (size selected for the 0.5-1 kb fragments) Cloned directionally, priming method: Oligo-dr. cDNA enrichment: >1k bp, Average insert size 1.8k bp. Priming sequence: 5'-GACTAGTTCTAGATCGGCGCGCC(T) 3'. Tissue contributed by: David Rowe. Library constructed by ResGen, Invitrogen Corp."

ORIGIN		Query Match		14.4%; Score 786.6; DB 6; Length 946;		Best Local Similarity		95.8%; Pred. No. 6.3e-186;		Matches 882; Conservative 0; Mismatches 31; Indels 8; Gaps 7;	
QY	3196	GCTGGGATATAGCTCGGTAGCAGAGCCCTCGCTACATGTCGAGGTCCTGGTTTTCAG	3255			Db	3	GCTGGGATATAGCTCGGTAGCAGAGCCCTCGCTACATGTCGAGGTCCTGGTTTTCAG	62		
QY	3256	TCCCCAGTACTGCAAAATAGAAAGAAAAACAATTGTCTTGATTAACATAAGGTTTAAAGCC	3315			Db	63	TCCCCAGTACTGCAAAATAGAAAGAAAAACAATTGTCTTGATTAACATAAGGTTTAAAGCC	122		
QY	3316	TCATAGTCAGTTCTAACTCAAATATATGCAATGATGATGTTCTTGTGCTTTTCTGTTTC	3375			Db	123	TCATAGTCAGTTCTAACTCAAATATATGCAATGATGATGTTCTTGTGCTTTTCTGTTTC	182		
QY	3376	TAAAAATTAATGGGCTTTATGGGTTTGTGTTTGTATGTTTGTATGTTTGTATATATTTGA	3435			Db	183	TAAAAATTAATGGGCTTTATGGGTTTGTGTTTGTATGTTTGTATGTTTGTATATTTGA	242		
QY	3436	GGTAGGGGATTTAGGAGGAGGCTCTCAGCTGGCCCTTGAGCTCATGCGAGTTTCTCGCTTC	3495			Db	243	GGTAGGGGATTTAGGAGGAGGCTCTCAGCTGGCCCTTGAGCTCATGCGAGTTTCTCGCTTC	302		
QY	3496	AACAGTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTT	3555			Db	303	AACAGTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTT	362		
QY	3556	GAAGGCAAAACCTGTTTATCAACCTAAAGACATTCGATGTCACCTTCAGTGGTGAATTT	3615			Db	363	GAAGGCAAAACCTGTTTATCAACCTAAAGACATTCGATGTCACCTTCAGTGGTGAATTT	422		

```

Qy 3616 TCTCCCTC-TTTTCTTTTCCCCAGAGCTGAGGACAAACCAGGCTTTGACATTGC 3674
Db 423 TCTCCCTCTTTTCTTTTCCCCAGAGCTGAGGACAAACCAGGCTTTGACATTGC 482
Qy 3675 TAGGCAAGCGCTTACACATGAGCTAAATCCCAACCCAGGCTTTTCTTTT 3734
Db 483 TAGGCAAGCGCTTACACATGAGCTAAATCCCAACCCAGGCTTTTCTTTT 541
Qy 3735 TAAAGACATGCTTATATAGTCTAGGCTGGCCCTTAAACTCATTAATAGCAAGGATGG 3794
Db 542 TAAAGACATGCTTATATAGTCTAGGCTGGCCCTTAAACTCATTAATAGCAAGGATGG 601
Qy 3795 CTTTGAAGATCTCTGCTCTCTTTTGGCTTGGCAATGGCAATGGCCATCATCCAG 3854
Db 602 CTTTGAAGATCTCTCTGCTCTCTTTTGGCTTGGCAATGGCCATCATCCAG 661
Qy 3855 ATTTTCTCTGTATCTAGGCTTTTAAAGATATTTTCAAGGACGCTCCCAAAATCCACAG 3914
Db 662 ATTTTCTCTGTATCTAGGCTTTTAAAGATATTTTCAAGGACGCTCCCAAAATCCACAG 721
Qy 3915 TGCACTACTTCTTCTGACTGGG-AAAGCGGGGTTCGTGCTTTTCAAGGACACACCAT 3973
Db 722 TGCACTACTTCTTCTGACTGGGAAAGCGGGTTCGTGCTTTTCAAGGACACACCAT 781
Qy 3974 TATGCCAGCTGCTGCTGCTTAAACATA-TTTGGTTTCTGTGCAAACTGTCCATTG 4032
Db 782 TATGCCAGCTGCTGCTGCTTAAACATA-TTTGGTTTCTGTGCAAACTGTCCATTG 841
Qy 4033 GAAGTTTCTGTGT-GTTGTAGTTTTCAGTTGAATG-TGGCTCTCTGTAGTTTCAAG--A 4088
Db 842 GAAGTTTCTGTGTGTGGGAAATTTTCAAGTTGAGTTGGCTCTCTGTAGTTTAAAGAA 901
Qy 4089 ATGGGTAGAAATCATAAGCA 4109
Db 902 AGGGTAAAGATCTTTAAGCA 922

RESULT 11
HSM805059 3941 bp mRNA linear HTC 22-SEP-2004
LOCUS Homo sapiens mRNA; cDNA DKFP6661186 (from clone DKFP6661186).
DEFINITION AL833746
VERSION AL833746.1 GI:21734396
KEYWORDS HTC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Bahr,A., Lauber,J., Mewes,H.W., Weil,B., Amid,C., Oeanger,A.,
Fobo,G., Han,M. and Wiemann,S.
The German cDNA Consortium
Direct Submission
Submitted (22-SEP-2004) MIPS, Ingolstaedter Landstr.1, D-85764
Neuherberg, GERMANY
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by Qiagen (Hilden/Germany) within the cDNA sequencing
consortium of the German Genome Project.
This clone (DKFP6661186) is available at the RZPD Deutsches
Reisourcenzentrum fuer Genomforschung GmbH in Berlin, Germany.
Please contact RZPD for ordering:
http://www.rzpd.de/cgi-bin/products/cl.cgi?CloneID=DKFP6661186
Further information about the clone and the sequencing project is
available at http://mips.gsf.de/projects/cdna/.
Location/Qualifiers
1. .3941
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="RZPD:DKFP6661186"
/db_xref="taxon:9606"
/clone="DKFP6661186"

FEATURES
source

```

```

/tissue_type="stomach"
/clone_lib="666 (synonym: hsto2). Vector pSport1; host
DH10B; sites NotI + SalI"
/dev_stage="adult"
/notes="hypothetical protein, N-terminus truncated"
1. .3941
/gene="DKFP6661186"
<2. .1225
/codon_start=1
/product="hypothetical protein"
/protein_id="CAH56240.1"
/db_xref="GI:52545825"
/translations="ASKGDLPTPDQDNGHSLLEBHSLLRMDPKHCRDNPOTFVPA
GDTGNQTSSTEEKIELLTFEVLALLESPPSCPSSEIPMEKFEGRISVVDVSY
LEGDLPIARPACSNKLDYILGASDLETSDEPEDWDEAEEDDFDSDSLSDS
DLQDPEGLHLWNSPCSVDPYNPONTATITQTAARIVPREPSSEKDLGSKSLDENS
QSGLPETPESHSSEEDWESSADEASLKNWSCNSDDPYNPLPKAPPQTSGENE
KGRDSTPSESISALISECHILSKVQLGSOSECPDPSVQRDLVLSGGKHTHVGRK
VTLEBVTYYISGDEDRKPWBEFARDGCRFQRIQETEDAIGYCLTFEHRERMFNR
LQTCFKGLNLVKQC"
ORIGIN
Query Match 14.4%; Score 786.4; DB 3; Length 3941;
Best Local Similarity 75.2%; Pred. No. 9.6e-186;
Matches 1108; Conservative 0; Mismatches 311; Indels 55; Gaps 8;
Qy 1361 GGCTAAACAAAGGCAAGAGTTACCCACCCCTGACCAAGATAATGGCTATCATAGCTTGA 1420
Db 1 GGCTAGCAGGGGCAAGATTATCCACCCCTGACCAAGATAATGGCTATCCACAGCTTGA 60
Qy 1421 GGAGGAACATAAATCTTCCCGATGGACCCACCAACATTCACAGATAACCCAGCAGGC 1480
Db 61 GGAGGAACACAGCTTCTCCGATGGATCCAAACACCTGCAGAGATAAACCCACACAGTT 120
Qy 1481 GGTGTCCCTCTGTCGACAGAGGCC-----GGAGCCCACTGAGAGAAAAACAGCA 1528
Db 121 TGTTCCTGCTGCTGGAGACATTCCTGGAAACACCCAGGAATCCACATGAAGAAAAATAGA 180
Qy 1529 ATTGGTGATTCAAGAAAGTTTCAC-----AGAGCCCCCAGGGAAGACGTCT 1573
Db 181 ATTATTAACTACAGAGTTTCCACTTGTGTTGGAAGAGAGAGAGCCCTCTGAGGCTGTCC 240
Qy 1574 GTTTTGTGAATTACCCGTGGAAAAAGAAAGTGTGAAGAGGACCAACATAATCAACTGACCT 1633
Db 241 ATCTAGTGAGATACCTATGGAAGAGAGCCCTGGAGAGGCGCGAATAGTGTAGTTGATTA 300
Qy 1634 CTGATAGAGGAGAGAGCCCTTCTGTTTCTACAGACCAAGTTTGTAGCAACAACTGAT 1693
Db 301 CTCATACCTAGAAAGGTGACCTTCCCATTTCTGCGCAGACCAAGCTTGTAGTAACAACTGAT 360
Qy 1694 AGATTATATTTTGGGAGGCGCCCTCAGTGACTTGTGAAGCCAGCTCTGATTCTGAAAGTGA 1753
Db 361 AGATTATATTTTGGGAGGTCATCCAGTGACCTGTGGAACAAAGTTCTGATCCAGAGGTGA 420
Qy 1754 GGAATGGGGCAGAAACCTTGAGGACGATGGCTTTGTATAGCATGGCTCCCTGTCTGAATC 1813
Db 421 GGATTGGGATGAGAAAGCTGAGGATGATGATGATGATGATGATGATGATGATGATGATGAT 480
Qy 1814 AGACGTGGAACAGAGCTCGGAAGCCCTTCACTTTTGGAACTCTTTTCCAGAGTGTAGATCC 1873
Db 481 AGACCTTGAACAGAGCCCTGGAAGGGCTTCACTTTTGGAACTCTTTCTGCAAGTGTAGATCC 540
Qy 1874 TTACAACCCCAAACTTTTACAGCCACGATTCAGAGCGCTGCCAGAAATTCGCCCCAGAGA 1933
Db 541 TTATAATCCCAAGACTTTTACAGCAAAATTCAGATCTGCTGCCAGAAATTTGTTCTGAAGA 600
Qy 1934 CCCATCAGATTTCAGGGACATCTCTGGTCTGCGACCTGTGGTGTAGGGA---GCTGTGAGCA 1990
Db 601 GCCTTCTGATTACAGAGAGGATTTGTCTGGCAAGTCTGATCTAGAGAAATTCCTCCAGATC 660
Qy 1991 GGACACCTTTCGGAGACCCCGACCATAGTTCCGGGGGAGGAAGATGACTGGGAACCGAG 2050

```





Db 421 AGAGTTTCCCTTAAACAAACACATGGCAGCTGCTCTTGGACATGTTTAAAGAAACAA 480  
Qy 2656 CTTGTATCTAGAGATGCAAGTTGATTAATTTTGGGTAAATGTCCTCATTAAGAAACACAA 2715  
Db CTTGTATCTAGAGATGCAAGTTGATTAATTTTGGGTAAATGTCCTCATTAAGAAACACAA 540  
Qy 2716 CTCGATAATGAAGATCTCTAATCTGTAAATCCCTCTCTTCCATTTAGTTGATGTGG 2775  
Db 541 CTCGATAATGAAGATCTCTAATCTGTAAATCCCTCTCTTCCATTTAGTTGATGTGG 600  
Qy 2776 GTTTTGTCTTTTGTAGAGGGTCTCACTGCATAATCTCTTTTGGCTAGGTCTATGTATGT 2835  
Db 601 GTTTTGTCTTTTGTAGAGGGTCTCACTGCATAATCTCTTTTGGCTAGGTCTATGTATGT 660  
Qy 2836 AGACAGGTTGGCCCTGAAGTTGGTAAATTCCTTGCCTCTGTTCCTGTGCACCATGCC 2895  
Db 661 AGACAGGTTGGCCCTGAAGTTGGTAAATTCCTTGCCTCTGTTCCTGTGCACCATGCC 720  
Qy 2896 AGCTGGAAGTGTTTTAAATATCTCTGCTTACCTGGGGTGAGAGTTTAAATTTGCACCT 2955  
Db 721 AGCTGGAAGTGTTTTAAATATCTCTGCTTACCTGGGGTGAGAGTTT-ATTTTGCACTT 779  
Qy 2956 TCAAAACGTTTCTTCTGGAGGCA 2978  
Db 780 TC-AAACGTTTCTTCTGGAGGCA 801

RESULT 13  
BI654191  
LOCUS 796 bp mRNA linear EST 12-SEP-2001  
DEFINITION 603280489P1 NCI\_QGAP\_Mam4 Mus musculus cDNA clone IMAGE:532446 5',  
mRNA sequence.

ACCESSION BI654191

VERSION 1 (bases 1 to 796)

KEYWORDS NIH-MGC http://mgi.nci.nih.gov/.

SOURCE National Institutes of Health, Mammalian Gene Collection (MGC)

ORGANISM Unpublished (1999)

REFERENCE Contact: Robert Strausberg, Ph.D.

AUTHORS Email: cgapbs-r@mail.nih.gov

TITLE Tissue Procurement: Lothar Hennighausen Ph.D., Priscilla Furth

JOURNAL Ph.D.

COMMENT cDNA Library Preparation: Life Technologies, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

Clone Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLM11823 row: a column: 23

High quality sequence stop: 785.

Location/Qualifiers

1. 796

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="NWR1"

/db\_xref="taxon:10090"

/clone="IMAGE:532446"

/tissue\_type="tumor, gross tissue"

/dev\_stage="5 months"

/lab\_host="DH10B"

/clone\_lib="NCI\_QGAP\_Mam4"

/note="Organ: mammary; Vector: pCMV-SPORT6; Site 1: SalI;

Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt.

Library constructed by Life Technologies. Investigators

providing samples: Lothar Hennighausen/Priscilla Furth,

NIH Reference for transgenic model: Li et al., Cell Growth

and Differentiation 7, 3-11 (1996)."

FEATURES

source

ORIGIN

Query Match 14.0%; Score 767.4; DB 4; Length 796;  
Best Local Similarity 99.1%; Pred. No. 4.1e-181;  
Matches 782; Conservative 0; Mismatches 6; Indels 1; Gaps 1;  
Qy 1 CGGTCTCTCGCTCCGCTCGAGCTTCGGGGTGGCGCTGCGGCCATTTTGGCTTCGC 60  
Db 1 CGGTCTCTCGCTCCGCTCGAGCTTCGGGGTGGCGCTGCGGCCATTTTGGCTTCGC 60  
Qy 61 TTCTTTTGGCGCTCGCTCGCACCAGCCACCTTTTCCGCTTGGCGCTTTTGGCGCTCCG 120  
Db 61 TTCTTTTGGCGCTCGCTCGCACCAGCCACCTTTTCCGCTTGGCGCTTTTGGCGCTCCG 120  
Qy 121 TGGGGGCAACGGAAACGGCCGCTCTCTCGCTCGCGCGCGCGAGAGGGTCTTCT 180  
Db 121 TGGGGGCAACGGAAACGGCCGCTCTCTCGCTCGCGCGCGCGAGAGGGTCTTCT 180  
Qy 181 CTATGTTGGAGCGATCTCACACGGCTTAGACGCTCTCTTCCCTAGCCGGATGGACCTA 240  
Db 181 CTATGTTGGAGCGATCTCACACGGCTTAGACGCTCTCTTCCCTAGCCGGATGGACCTA 240  
Qy 241 ACCCGGTCGCGACCGCTTGGCGGGCTCTTGGGCCGCTCCGCTGCAGCATCTCGTTGCGGA 300  
Db 241 ACCCGGTCGCGACCGCTTGGCGGGCTCTTGGGCCGCTCCGCTGCAGCATCTCGTTGCGGA 300  
Qy 301 AGCGCGCTCTCTTGGGCTCTCTTGGCGCGCGGGGAATCGAGTACCCACCTCCG 360  
Db 301 AGCGCGCTCTCTTGGGCTCTCTTGGCGCGCGGGGAATCGAGTACCCACCTCCG 360  
Qy 361 TGGCTGGGCAAGCGGAGACTGTGTAGACCTCGGATCCAGCTGCGCTGAGCGCGCTGAG 420  
Db 361 TGGCTGGGCAAGCGGAGACTGTGTAGACCTCGGATCCAGCTGCGCTGAGCGCGCTGAG 420  
Qy 421 CTCTGTCTCTCTCTGTCTGAGAACCGCCCAAGAAAGAGATGAGACAGGAAACGACA 480  
Db 421 CTCTGTCTCTCTCTGTCTGAGAACCGCCCAAGAAAGAGATGAGACAGGAAACGACA 480  
Qy 481 GGGCCGGAAGCGGCTTGGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCT 540  
Db 481 GGGCCGGAAGCGGCTTGGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCT 540  
Qy 541 GATCGACGCTGCTCTTCCGAGTTCGCGCGCTTCTCTCGACAAATCCCGGAACT 600  
Db 541 GATCGACGCTGCTCTTCCGAGTTCGCGCGCTTCTCTCGACAAATCCCGGAACT 600  
Qy 601 CCGCTTGGCCGAGCTGCTGAGCAGGCTGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAG 660  
Db 601 CCGCTTGGCCGAGCTGCTGAGCAGGCTGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAG 660  
Qy 661 TGCTTCCCTAGCTTATTCAGAAAGCTGCTGCTTTTGGAGCCAGCTTTCCGGGGGCTGATTC 720  
Db 661 TGCTTCCCTAGCTTATTCAGAAAGCTGCTGCTTTTGGAGCCAGCTTTTCCGGGGGCTGATTC 720  
Qy 721 CTACCAGATGCTAGATTTTGGCGCAAGTTACAGCGCCCTGAGAGCTTCAGAGGAGCGGG 780  
Db 721 CTACCAGATGCTAGATTTTGGCGCAAGTTTTCGCAAGTTTACAGCGCCCTGAGAGGAGCGGG 779  
Qy 781 AGGAATCTG 789  
Db 780 AGGAATCTG 788

RESULT 14

CF746885

LOCUS

DEFINITION

UI-M-H80-cmb-g-24-0-UI.r1 NIH\_BMAP\_HE0 Mus musculus CDNA clone

IMAGE:30627143 5', mRNA sequence.

ACCESSION CF746885

VERSION

CF746885.1

KEYWORDS

EST.

SOURCE

Mus musculus (house mouse)

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

CF746885 854 bp mRNA linear EST 10-OCT-2003

UI-M-H80-cmb-g-24-0-UI.r1 NIH\_BMAP\_HE0 Mus musculus CDNA clone

IMAGE:30627143 5', mRNA sequence.

ACCESSION CF746885

VERSION

CF746885.1

KEYWORDS

EST.

SOURCE

Mus musculus (house mouse)

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;



/note="Organ: placenta; Vector: pExpress-1; Site:1: EcoRV; Site 2: NotI; RNA obtained from three placentas from female C57/BL6 mouse at 16 days pregnancy. Tissues were snap-frozen and kept at -80C for two days before RNA extraction and purification (Tri-reagent method). cDNA was primed using oligo-dT primer: 5'-pGACTAGTCTAGATCGAGCGCCGCC(T)25-3' and cloned into the EcoRV/NotI sites of pExpress-1. Size-selection >0.75kb resulted in an average insert size of 1.1 kb. This primary, nanocount library is normalized to Cot5 (non-normalized primary library is NIH MGC-223) and was constructed by Express Genomics (Frederick, MD)."

Search completed: September 16, 2005, 08:56:22  
Job time : 16341.4 secs

ORIGIN

Query Match	13.9%	Score 759.6;	DB 7;	Length 835;
Best Local Similarity	99.2%	Pred. No. 3.8e-179;		
Matches 784;	Conservative 0;	Mismatches 4;	Indels 2;	Gaps 2;
Qy	3992	CAAACTTAAACATATTTGGTTTCTGTGCAAACTGTCCATTTGGAAAGGTTTCTGTGTGTGT 4051		
Db	3	CCAACTTAAACATATTTGGTTTCTGTGCAAACTGTCCATTTGGAAAGGTTTCTGTGTGTGT 62		
Qy	4052	AGTTTTCAGTTCGAATGTGCTCTCTGTAGCTTTCAAGATGGTAGAATCATTAAGCACT 4111		
Db	63	AGTTTTCAGTTCGAATGTGCTCTCTGTAGCTTTCAAGATGGTAGAATCATTAAGCACT 122		
Qy	4112	CTTAAGTAATCATTTGCAATTTAGACATTTTTTTTTTTTAACTAGGGGGTATTTGGGAGAC 4171		
Db	123	CTTAAGTAATCATTTGCAATTTAGACATTTTTTTTTTTTAACTAGGGGGTATTTGGGAGAC 181		
Qy	4172	ATGTCAGATATTCATTTTGTGTTTATGTGCTTAAACCAGTGTTACTTACACCTGCCT 4231		
Db	182	ATGTCAGATATTCATTTTGTGTTTATGTGCTTAAACCAGTGTTACTTACACCTGCCT 241		
Qy	4232	CAGCACTGGCACTTTTCAAACTGCTTGGAGACACTGTAACTTGGATGGTGAGTTCT 4291		
Db	242	CAGCACTGGCACTTTTCAAACTGCTTGGAGACACTGTAACTTGGATGGTGAGTTCT 301		
Qy	4292	GGGTTTTTCATGTGTAATAATGTCAACTACAAAGGTCAATATCCAGGTTTGTGTGTCTAC 4351		
Db	302	GGGTTTTTCATGTGTAATAATGTCAACTACAAAGGTCAATATCCAGGTTTGTGTGTCTAC 361		
Qy	4352	CTTATGTAGAGAGAACAAAGCAGAAAGGGCAGATAGAGCAGCCAGAAAGTGCTAGTGT- 4410		
Db	362	CTTATGTAGAGAGAACAAAGCAGAAAGGGCAGATAGAGCAGCCAGAAAGTGCTAGTGT 421		
Qy	4411	CCCCCCCCAAGCGTCTCTAGATAGTGTGGATCAGGTGCTTGGTTTGTTCAGTTA 4470		
Db	422	CCCCCCCCAAGCGTCTCTAGATAGTGTGGATCAGGTGCTTGGTTTGTTCAGTTA 481		
Qy	4471	CGGTCACTGTGTGCAATGGTTGAGGTAATGTGTTTACCAGTTCTTTAGTGGTTACATGCA 4530		
Db	482	CGGTCACTGTGTGCAATGGTTGAGGTAATGTGTTTACCAGTTCTTTAGTGGTTACATGCA 541		
Qy	4531	CAAAAGAGAGAACCTCTGAGTCGGGTGFGGATGAGCTTCCAGACCTGGCAGGGTAAAC 4590		
Db	542	CAAAAGAGAGAACCTCTGAGTCGGGTGFGGATGAGCTTCCAGACCTGGCAGGGTAAAC 601		
Qy	4591	TACCTCAGTTTATATCTCCCTGGTTATTTCCGTTTGTATGATCTAAGTCTGCCTCAG 4650		
Db	602	TACCTCAGTTTATATCTCCCTGGTTATTTCCGTTTGTATGATCTAAGTCTGCCTCAG 661		
Qy	4651	TGGTGATGATGTTTATCCACACACAGGTTAGTAAGAGTGACGACCCAGAAACGTTGGTC 4710		
Db	662	TGGTGATGATGTTTATCCACACACAGGTTAGTAAGAGTGACGACCCAGAAACGTTGGTC 721		
Qy	4711	TTATTTTGGAGACCCCACTTCTGTGTAATTTTATGCACTTGCCTTTAGTGAACCTCCAGA 4770		
Db	722	TTATTTTGGAGACCCCACTTCTGTGTAATTTTATGCACTTGCCTTTAGTGAACCTCCAGA 781		
Qy	4771	GTGCATTAAA 4780		
Db	782	GTGCATTAAA 791		

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 15, 2005, 16:13:49 ; Search time 68.7612 Seconds  
(without alignments)  
3926.035 Million cell updates/sec

Title: US-10-650-482-4

Perfect score: 3710

Sequence: 1 METGTHARKRPGRLGSWF.....MFNRLRIESKDLLYSNVKX 698

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : A\_Geneseq\_16Dec04.\*

1: Geneseqp1980s.\*

2: Geneseqp1990s.\*

3: Geneseqp2000s.\*

4: Geneseqp2001s.\*

5: Geneseqp2002s.\*

6: Geneseqp2003as.\*

7: Geneseqp2003bs.\*

8: Geneseqp2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3710	100.0	698	8	Adq88460 Mouse GAD
2	2223	59.9	713	8	Adq88458 Human GAD
3	2213	59.6	713	4	Aab95876 Human pro
4	2157.5	58.2	707	3	Aab53401 Human col
5	547	14.7	153	4	Aam25448 Human pro
6	204	5.5	657	5	Abb57373 Mouse isc
7	201	5.4	657	2	Aay41102 Rat Myd11
8	191.5	5.2	674	2	Aaw99891 Human gro
9	189.5	5.1	674	3	Aay84366 Amino aci
10	189.5	5.1	674	7	Ad54748 Human pro
11	189.5	5.1	674	7	Add45119 Human pro
12	189.5	5.1	674	7	Ad54744 Human pro
13	186.5	5.0	674	7	Aab92888 Human pro
14	186.5	5.0	674	8	Adr14089 Human NF-
15	182.5	4.9	578	2	Aaw79958 Human pro
16	182.5	4.9	578	2	Aay06514 Human pro
17	182.5	4.9	578	2	Aay41104 Human pro
18	171.5	4.6	590	2	Aay41101 Rat gacd3
19	149.5	4.0	923	6	Abu54617 Human NOV
20	149	4.0	1365	8	Adq39691 Human myo
21	148.5	4.0	676	4	Abb62247 Drosophil
22	147.5	4.0	725	4	Aab95238 Human pro
23	147.5	4.0	835	5	Abb05681 Human dif
24	147.5	4.0	835	5	Adr14035 Human NF-
25	147.5	4.0	835	8	Abm80164 Tumour-as

26	144	3.9	867	5	ABP35633	Abp35633 Fungal ZB
27	141	3.8	917	6	ADA13333	Ada13333 Human int
28	141	3.8	1198	8	ADR09843	Adr09843 Human pro
29	139.5	3.8	883	6	ABP72604	Abp72604 Rat brain
30	139.5	3.8	883	6	ADK67782	Adk67782 Rat glyco
31	139.5	3.8	990	5	AAE22762	Ade22762 Rat PGC-3
32	138.5	3.7	1454	8	ABM84777	Abm84777 Human dia
33	138	3.7	1572	5	ABB97562	Abb97562 Novel hum
34	138	3.7	1572	6	ABP98339	Abp98339 Amino aci
35	137	3.7	1088	8	ADQ65665	Adq65665 Novel hum
36	136.5	3.7	883	6	ABP72603	Abp72603 Rat mutan
37	136.5	3.7	883	6	ADK67779	Adk67779 Rat brain
38	136.5	3.7	1234	8	ADQ39694	Adq39694 Human myo
39	136.5	3.7	1366	8	ADQ39692	Adq39692 Human myo
40	136.5	3.7	1411	8	ADQ39693	Adq39693 Human dia
41	136	3.7	826	7	ABM84423	Abm84423 Human dia
42	133.5	3.6	883	7	ADE62581	Ade62581 Rat Prote
43	132.5	3.6	995	8	ADR98899	Adr98899 Lung spec
44	132	3.6	1715	4	ABBS8089	Abbs8089 Drosophil
45	131.5	3.5	1662	4	AAU32140	Aau32140 Novel hum
46	131.5	3.5	1855	7	AD73118	Ad73118 Human cel
47	131.5	3.5	1902	7	ADH77759	Adh77759 Human Als
48	131.5	3.5	3708	7	ABM85423	Abm85423 Human pro
49	131.5	3.5	4127	6	ABR82115	Abm82115 Human ALM
50	130	3.5	1596	5	AAE20271	Aae20271 Human lun
51	129	3.5	817	4	AAU31937	Aau31937 Novel hum
52	129	3.5	930	6	ABB99465	Abb99465 Amino aci
53	127.5	3.4	406	5	ABG93152	Abg93152 S. cerevi
54	127.5	3.4	1003	8	AD73317	Ad73317 Tobacco 3
55	127.5	3.4	1078	8	ADQ66807	Adq66807 Novel hum
56	127.5	3.4	1281	8	ADR09773	Adr09773 Human pro
57	127.5	3.4	3913	7	ADK40938	Adk40938 Novel hum
58	127.5	3.4	3913	8	ADRI5659	Adri5659 Kinase 51
59	127	3.4	885	7	ADP28591	Adp28591 Neural Ig
60	127	3.4	920	8	ADQ20144	Adq20144 Human sof
61	127	3.4	996	5	AAE22760	Aae22760 Human PGC
62	127	3.4	1002	5	AAE22759	Aae22759 Human PGC
63	127	3.4	1010	7	ADP28580	Adp28580 Neurologi
64	127	3.4	1023	5	AAE22761	Aae22761 Human PGC
65	127	3.4	1028	5	AD128022	Ad128022 ECMCAD pr
66	127	3.4	1028	7	ADP28577	Adp28577 Neural Ig
67	127	3.4	1028	8	ADSI0519	Adsi0519 Human the
68	127	3.4	1056	7	ADD18754	Add18754 Human dis
69	127	3.4	1454	7	ADD93813	Add93813 Human apo
70	127	3.4	1503	2	AAW48845	Aaw48845 Human rec
71	127	3.4	1503	2	ADD93815	Add93815 Human HJ0
72	127	3.4	1503	2	ADP55733	Adp55733 Human PRO
73	126	3.4	237	2	AAE6187	Aae6187 Human bla
74	126	3.4	817	7	ADE54704	Ade54704 Rat Prote
75	126	3.4	2759	6	AAO16418	Aao16418 Human nuc
76	125.5	3.4	1197	4	AAU95146	Aau95146 Human pro
77	125.5	3.4	1251	5	AAU98889	Aau98889 Human pro
78	125.5	3.4	1270	5	ABP69291	Abp69291 Human pol
79	125.5	3.4	1316	7	ADF28764	Adf28764 Human deu
80	125	3.4	2781	3	AAE57453	Aae57453 Human tra
81	125	3.4	2781	8	ADP42724	Adp42724 Human BPT
82	125	3.4	2781	8	ADQ00984	Adq00984 Human hom
83	125	3.4	2781	8	ADQ18653	Adq18653 Human sof
84	125	3.4	2907	3	AAE57452	Aae57452 Human tra
85	124.5	3.4	1354	4	ABB62359	Abb62359 Drosophil
86	124.5	3.4	1532	4	AAU71956	Aau71956 Human bon
87	124.5	3.4	1780	2	AAW53863	Aaw53863 Human gra
88	124.5	3.4	1780	2	AAW53863	Aaw53863 Human gra
89	124	3.3	605	3	AAE57950	Aae57950 Human tra
90	124	3.3	640	4	ABG11513	Abg11513 Novel hum
91	124	3.3	1795	4	ABG21018	Abg21018 Novel hum
92	123.5	3.3	582	4	AAW51816	Aaw51816 Murine ap
93	123.5	3.3	718	4	AAW80140	Aaw80140 Human pro
94	123.5	3.3	4274	4	ABG00972	Abg00972 Novel hum
95	123.5	3.3	4386	4	ABG07375	Abg07375 Novel hum
96	123.5	3.3	4397	4	ABG21944	Abg21944 Novel hum
97	123	3.3	1023	7	ABM04438	Abm04438 Human pro
98	123	3.3	1142	2	AAW81546	Aaw81546 Tumour re



99	123	3.3	1142	3	AAY43876	Aay43876 Amino aci
100	123	3.3	1142	6	ABU08929	Abu08929 Human tum
ALIGNMENTS						
RESULT 1						
ID	ADQ88460	ADQ88460 standard; protein; 698 AA.				
AC	ADQ88460;					
DT	07-OCT-2004	(first entry)				
DE	Mouse	GADD34-like (GADD34L) protein.				
XX						
KW	Oxidative stress; neuronal ischaemia; heart ischaemia; renal damage;					
KW	autoimmune disease; neurodegenerative disorder; therapy; GADD34-like;					
KW	GADD34L; eIF2alpha-specific regulatory subunit; phosphatase; mouse.					
XX						
OS	Mus sp.					
XX						
FN	US2004142345-A1.					
PD	22-JUL-2004.					
XX						
PF	28-AUG-2003; 2003US-00650482.					
XX						
PR	06-SEP-2002; 2002US-0408679P.					
PA	(ROND/) RON D.					
PA	(JOUS/) JOUSSE C.					
XX						
PI	Ron D, Jousse C;					
XX						
DR	WPI; 2004-552556/53.					
DR	N-PSDB; ADQ88459.					
XX						
PT	Screening test substances for preventing or treating disease involving					
PT	oxidative stress, by testing test substances for its ability to inhibit					
PT	activity of GADD34L and identifying test substance that inhibits activity					
PT	of GADD34L.					
XX						
PS	Disclosure; SEQ ID NO 4; 30pp; English.					
XX						
CC	The present invention relates to a method of screening several test					
CC	substances for preventing or treating diseases involving oxidative stress					
CC	such as neuronal ischaemia, heart ischaemia, renal damage induced by					
CC	ischaemia or toxins, autoimmune diseases and neurodegenerative disorders.					
CC	The method involves testing the test substances for its ability to					
CC	inhibit the activity of GADD34-like (GADD34L), also referred to as					
CC	eIF2alpha-specific regulatory subunit of phosphatase, and identifying the					
CC	test substance which inhibits the activity of GADD34L. The present					
CC	sequence is the mouse GADD34L protein.					
XX						
SQ	Sequence 698 AA;					
Query Match						
Best Local Similarity 100.0%; Score 3710; DB 8; Length 698;						
Matches 698; Conservative 0; Mismatches 0; Indels 0; Gaps 0;						
QY	1	METGTHARKPGPRLGFWRLPFLRRSHACSSPPPPSSRNQNSALPERRTRYWTKL	60			
Db	1	METGTHARKPGPRLGFWRLPFLRRSHACSSPPPPSSRNQNSALPERRTRYWTKL	60			
QY	61	LSQLLALLPSLFQKLLLSQSGGLIPTRWLDFAASYSALRASGRSDAPTQKSLSY	120			
Db	61	LSQLLALLPSLFQKLLLSQSGGLIPTRWLDFAASYSALRASGRSDAPTQKSLSY	120			
QY	121	TAAGLFAKTRVVSTLALRGCTPVAVLRLVLEVKLKAQERALDSAAPTFLLEQQLWGVEL	180			
Db	121	TAAGLFAKTRVVSTLALRGCTPVAVLRLVLEVKLKAQERALDSAAPTFLLEQQLWGVEL	180			

QY	181	LPSSLQAGLVSHRELDSSSSGPLSVQSGNFKVQSYLLNPSYLDYLPGLRCQSSAGG	240			
Db	181	LPSSLQAGLVSHRELDSSSSGPLSVQSGNFKVQSYLLNPSYLDYLPGLRCQSSAGG	240			
QY	241	QFVGFRLLTPESCYLESDGCHPQPLRAEMSAWARRCPPLSTEGELPEIHRRMRWLVLQ	300			
Db	241	QFVGFRLLTPESCYLESDGCHPQPLRAEMSAWARRCPPLSTEGELPEIHRRMRWLVLQ	300			
QY	301	PNQGDLPDLDQDNGYHSLSEEHNLRLMDPQHCTDNPAQAVSPAADRPEPEKPELVQ	360			
Db	301	PNQGDLPDLDQDNGYHSLSEEHNLRLMDPQHCTDNPAQAVSPAADRPEPEKPELVQ	360			
QY	361	EVSQSPQSSLFCELPVEKECEDHTNATDLSDRGESLPVSTRPVCNKLIDYILGGAPS	420			
Db	361	EVSQSPQSSLFCELPVEKECEDHTNATDLSDRGESLPVSTRPVCNKLIDYILGGAPS	420			
QY	421	DLEASDSESDWGEPEDDGFDSDGSLSESDEQDSGLHLWNSFHSVDPYKPNFTAT	480			
Db	421	DLEASDSESDWGEPEDDGFDSDGSLSESDEQDSGLHLWNSFHSVDPYKPNFTAT	480			
QY	481	IQTAAIARPDPSDGTSGTSGVSGCOEGLPPTPDHSGSEEDWEPSPADEANLKLW	540			
Db	481	IQTAAIARPDPSDGTSGTSGVSGCOEGLPPTPDHSGSEEDWEPSPADEANLKLW	540			
QY	541	NSFCHSEDPYNLLNFKAPQPSGKWKGRQDSKASSEVTVAFSGHHTLLSCKAQLLESQE	600			
Db	541	NSFCHSEDPYNLLNFKAPQPSGKWKGRQDSKASSEVTVAFSGHHTLLSCKAQLLESQE	600			
QY	601	DNCPGCGLGEALAGERYTHIKKKVTFLEEVTEYYISGDEDRKGPWSEFARDGCRFKRI	660			
Db	601	DNCPGCGLGEALAGERYTHIKKKVTFLEEVTEYYISGDEDRKGPWSEFARDGCRFKRI	660			
QY	661	QETEVAIGCYCLAFEHREKMFNRLRIESKOLLYSNVKK	698			
Db	661	QETEVAIGCYCLAFEHREKMFNRLRIESKOLLYSNVKK	698			
RESULT 2						
ID	ADQ88458	ADQ88458 standard; protein; 713 AA.				
AC	ADQ88458;					
XX	07-OCT-2004	(first entry)				
XX	Human	GADD34-like (GADD34L) protein.				
XX	Oxidative stress; neuronal ischaemia; heart ischaemia; renal damage;					
XX	autoimmune disease; neurodegenerative disorder; therapy; GADD34-like;					
XX	GADD34L; eIF2alpha-specific regulatory subunit; phosphatase; human.					
XX	Homo sapiens.					
XX	US2004142345-A1.					
XX	22-JUL-2004.					
XX	28-AUG-2003; 2003US-00650482.					
XX						
PR	06-SEP-2002; 2002US-0408679P.					
XX						
PA	(ROND/) RON D.					
PA	(JOUS/) JOUSSE C.					
XX						
PI	Ron D, Jousse C;					
XX						
DR	WPI; 2004-552556/53.					
DR	N-PSDB; ADQ88457.					
XX						
Screening test substances for preventing or treating disease involving						
oxidative stress, by testing test substances for its ability to inhibit						
activity of GADD34L and identifying test substance that inhibits activity						
of GADD34L.						

QY	181	LPSSLQAGLVSHRELDSSSGPLSVQSLGNFKVVS YLLNPSYLDYLPQLGLRCOSSAGGG	240
DB	181	LPSSLQAGLVSHRELDSSSGPLSVQSLGNFKVVS YLLNPSYLDYLPQLGLRCOSSAGGG	240
QY	241	QVGFRTLTPESCYSLEDGCHPQPLRAEMGATARRCPPLSTEGLPPIHHRMRWLVLQ	300
DB	241	QVGFRTLTPESCYSLEDGCHPQPLRAEMGATARRCPPLSTEGLPPIHHRMRWLVLQ	300
QY	301	PNQGDLPITLDQDNGYHSLFEEHNLLRMDFQHCCTDNPAQAVSPAADRPETKKEPELVQ	360
DB	301	PNQGDLPITLDQDNGYHSLFEEHNLLRMDFQHCCTDNPAQAVSPAADRPETKKEPELVQ	360
QY	361	EVSQSPQSSILFCELPVEKECEEDHTNATDLSRGESLPVSTRVPCSNKLIIDYILGAPS	420
DB	361	EVSQSPQSSILFCELPVEKECEEDHTNATDLSRGESLPVSTRVPCSNKLIIDYILGAPS	420
QY	421	DLEASSDSESEDWGEPEDDGFDSDGSLSDVEQDSEGLHLMNSFHSVDPYPQNFAT	480
DB	421	DLEASSDSESEDWGEPEDDGFDSDGSLSDVEQDSEGLHLMNSFHSVDPYPQNFAT	480
QY	481	IQTAAIAPRDPDSGTSWSGCGVSCQSGPLPETPDHSSGDEDDWEPSSADEAKLW	540
DB	481	IQTAAIAPRDPDSGTSWSGCGVSCQSGPLPETPDHSSGDEDDWEPSSADEAKLW	540
QY	541	NSFCHSEDVYNLLNFKAPFQPGKQWKGRQDSKASSEVTVAFSGHHTLLSCKAQLLESQE	600
DB	541	NSFCHSEDVYNLLNFKAPFQPGKQWKGRQDSKASSEVTVAFSGHHTLLSCKAQLLESQE	600
QY	601	DNCPGCGLGEALAGERYTHIKRKVTFLEEVTEYIISGDEDRKGPWEEFARDGCRFQRI	660
DB	601	DNCPGCGLGEALAGERYTHIKRKVTFLEEVTEYIISGDEDRKGPWEEFARDGCRFQRI	660
QY	661	QETEVAICYCLAFEHREKMFNRLIESKDLLLSYNVKK	698
DB	661	QETEVAICYCLAFEHREKMFNRLIESKDLLLSYNVKK	698

RESULT 2			
ADQ88458			
ID	ADQ88458 standard; protein; 713 AA.		
XX			
AC	ADQ88458;		
XX			
DT	07-OCT-2004 (first entry)		
XX			
DE	Human GADD34-like (GADD34L) protein.		
XX			
KW	Oxidative stress; neuronal ischaemia; heart ischaemia; renal damage;		
KW	autoimmune disease; neurodegenerative disorder; therapy; GADD34-like;		
KW	GADD34L; eIF2alpha-specific regulatory subunit; phosphatase; human.		
XX			
OS	Homo sapiens.		
XX			
FN	US2004142345-A1.		
XX			
PD	22-JUL-2004.		
XX			
PF	28-AUG-2003; 2003US-00650482.		
XX			
PR	06-SEP-2002; 2002US-0408679P.		
XX			
PA	(ROND/) RON D.		
PA	(JOUS/) JOUSSE C.		
XX			
PI	Ron D, Jousse C;		
XX			
DR	WPI; 2004-552556/53.		
DR	N-PSDB; ADQ88457.		
XX			
PT	Screening test substances for preventing or treating disease involving		
PT	oxidative stress, by testing test substances for its ability to inhibit		
PT	activity of GADD34L and identifying test substance that inhibits activity		

```
PT of GADD34L.
PS Disclosure; SEQ ID NO 2; 30pp; English.
XX
XX The present invention relates to a method of screening several test
CC substances for preventing or treating diseases involving oxidative stress
CC such as neuronal ischaemia, heart ischaemia, renal damage induced by
CC ischaemia or toxins, autoimmune diseases and neurodegenerative disorders.
CC The method involves testing the test substances for its ability to
CC inhibit the activity of GADD34-like (GADD34L), also referred to as
CC eIF2alpha-specific regulatory subunit of phosphatase, and identifying the
CC test substance which inhibits the activity of GADD34L. the present
CC sequence is the human GADD34L protein.
XX
XX Sequence 713 AA;
SQ
    Query Match      59.9%; Score 2223; DB 8; Length 713;
    Best Local Similarity 65.0%; Pred. No. 5.6e-130;
    Matches 456; Conservative 53; Mismatches 174; Indels 18; Gaps 7;
Qy 1 METGTTHARKRPGRLGFWRLPFL-RRSHACSEFFPPSSRONPGN-----SALPERRT 54
Db 1 MEPTGSGRKLGRAGRFWPPPPRRSQAGSKFPTPLGPNSEGNFTLLSSAQPTRV 60
Qy 55 RYWTKLLSQLALLPSLFQKLLWSQLSGGLIPTRWLDFAAYSALRASRGRESADPTV 114
Db 61 SYWTKLLSQLLAPLPGLLQKVLWSQLFGGFWPTRWLDFAGVYSALRALKGREKPAFTA 120
Qy 115 QKSLSYTAAGLFAKTRVVSTLALARGTTPAVLVRLVKLKAQERALDSAAPTFLEQQ 174
Db 121 QKSLSSQLQD--SSDPSVTSPLDWLEEGTHWQYSPDKLKLKAGSALDPAQAFLLEQQ 179
Qy 175 LMGVELLPSSIQAGLVSHREDDSSSGPLSVQSLGNFKVSVYLLNPSVLDYLPOLGLRCQ 234
Db 180 LMGVELLPSSLQSLRYNRELSSPSPGLNQLQINDFNSVSVYLLNPSVLDYLPOLGLRCQ 239
Qy 235 SSAGGQFVGPRTTLTPESCYLEDCGCHPQPLRAEMSAATAMRCCPLSTEGELPEIHRMR 294
Db 240 NSDGNSEVVGFTLTPESSCLREDHCHPQPLSABLIPASWQCFPLSTEGELPEIHRMR 299
Qy 295 WLVEL-QNQODI.PTLDODNGYHSLREHNLBWDQCHTNDPAQVSPAADRP----E 349
Db 300 RLEFLQQAANKQDLPFTPDQDNGYHSLREHNSLLRMDPKHCRDNPQTQVPAAGDIPGNTQE 359
Qy 350 PTEKKPELVIOEV-----SQSPQSGSLFCELPVKECEEDHTNATDLSDRGESLPVSTRP 404
Db 360 STEEKIELLTTEVPLALEERSPSGCSSEIPEMEKEPGEGRISVVDYSYLEGDLPIGARP 419
Qy 405 VCSNKLIDYILGGAPSDLEASDSESDWGEPEDDGFDPSDGLSESDVEQDSGLHLWN 464
Db 420 ACSNKLIDYILGGASDLETSSDPEGEDWDEAEDDGFSDSLSLSDLEQDPEGLHLWN 479
Qy 465 SFHSDVPKPNQFTATTQTAARIAPRDSDSGTWSGCGV-GSCQGPLPETHDSGE 523
Db 480 SFCSDVPKPNQFTATTQTAARIPEPSDEKDLGSKSDLENSQSGSLPETPEHSGE 539
Qy 524 EDDPEPSADEENLKLWNSFCHSDPYNLLNFKAPPQPSGKNWKGRODSKASSVTVAFS 583
Db 540 EDDMESSADEAESLKLWNSFCNSDDPYNLLNFKAPPQPSGKNWKGRODSKTPSESIVAS 599
Qy 584 GHHTLLSCKAQLLESQEDNCPCGGLGEALAGERYTHIKRKKVTFLEEVTEYIISGDEDRK 643
Db 600 ECHTLLSCKVQLLSQESCECPDSVQDVLSSGRHHTVKKVTFLEEVTEYIISGDEDRK 659
Qy 644 GPWEEFARDGCRFQRIQETEVAGYCLAFPHREKMFNRLR 684
Db 660 GPWEEFARDGCRFQRIQETEDAIGYCLTFPHREMFNRLQ 700
RESULT 3
AAB95876
ID AAB95876 standard; protein; 713 AA.
XX
```

115	QKSLSYTAAGLFAKTRTVVSTLTALAGCTPVAIVLRLVLEVKLKAQERALDSAAFTLLEQ	174
121	QKSLSLQID--SDPSVTSPLDMEEGTHWQYSPDCLKLELKAQSGALSDPAAQAFLLQ	179
175	LWGVELLPSSLOAGLVSHRELDSSSGPLSVQSLGNFKVSYLLNPYSYLVLPOLGLRCQ	234
180	LWGVELLPSSLOASRLYSNRELSSPSGPIQRTDDFSVSYLLNPYSYLVLPOLGLRCQ	239
235	SSAGGQFVGPRTLTPESCYSLEDGCHPQPIRAEMSATAMRRCPLSTEGLEPIHHRMR	294
240	NSDGNSEVVGFQTLPESSCLREDHCHPQLSAELIPASWGCPPLSTEGLEPIHHLRMK	299
295	WLVLPL-OPNQGDPLTLDODNGYHLEBEHNLLRMDPOHCTDNPQAQVSPAADRP----	349
300	RLEFLOAQSGKGLDTPPDQNGYHLEBEHLLRMDPKHCRDNPQTPVPAAGIPGNTOC	359
350	PTEKPELVIQEV----SOSPOGSSLCELPEVKECEEDHNTATLSDRGESLPIVSTRP	404
360	STEEKLELTTVEPLALREESPECCPSEIPEKBPGEGRISVDVYSYLEGOLPISAP	419
405	VCSNKLIIDYILGAPSDLEASDSSESDWGEPEDDGFDSDGSLSESDVQDSGEGHLWN	464
420	ACSNKLIIDYILGASSDLETSDPEGEDWDEAEDDGFDSDSLSDSLDQDEGLHLWN	479
465	SFHSVDYPKQNFATTIQTAAIARPDSDSGTSHWSGCGV-GSCQGGPIPETPDHSSG	523
480	SFCSVDYPNQNFATTIQTAAIRVPEEPSDEKDLGKSLDLENSQSGLSPETPEHSSG	539
524	EDDWPSADEANLKLWNSPCHSEDPYLLNFKAPQPSGKNWKGRODSKASSEVTVAS	583
540	EDDWSSADEASLKLWNSFCNSDDPYNPLNFKAPQTSGENEGKCKDSKTPGESIVAS	599
584	GHTLLSCKAQLLESOENCPGCLGEALAGERYTHIKKKVTPLEEVYIYISGDEDRK	643
600	ECHTLLSCKQLLGQSECECPDSVQDVLSGRHTHVKKVTFLEEVYIYISGDEDRK	659
644	GPWEEFARGCFQKRIQETEVAIQYCLAFEHREKMFNRLR	684
660	GPWEEFARGCFQKRIQETDAIGYCLTFHREMFNRLQ	700

## RESULT 4

AAB53401	
ID	AAB53401 standard; protein; 707 AA.
XX	XX
AC	AAB53401;
XX	XX
DT	09-MAR-2001 (first entry)
XX	XX
DE	Human colon cancer antigen protein sequence SEQ ID NO:941.
XX	XX
KW	Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW	identification; cytostatic; cardioactive; neuroprotective; vulnerary;
KW	immunomodulatory; muscular; gynaecological; gastrointestinal;
KW	nephrotropic; antineoplastic; antibacterial; gene therapy; wound;
KW	neural disorder; immune system disorder; muscular disorder;
KW	reproductive disorder; gastrointestinal disorder; renal disorder;
KW	infectious disease; cardiovascular disorder.
XX	XX
OS	Homo sapiens.
PN	WQ20005351-A1.
XX	XX
PD	21-SEP-2000.
PX	XX
PF	XX
PI	08-MAR-2000; 2000WO-US005883.
XX	XX
PR	12-MAR-1999; 99US-0124270P.
PX	XX
PA	(HUMA-) HUMAN GENOME SCI INC.
XX	XX
PI	Rosen CA, Ruben SM;
XX	XX

Db 606 LGSQSECPDSVQRDVLSSGRHTVKKVTFLEBVTYYISGDDRGKWPWFARDGCR 665  
Qy 656 FOKRIQETVAIGYCLAFEHREKMFNRLR 684  
Db 666 FOKRIQETDAIGYCLTFEHREMFNRLQ 694

RESULT 5  
ID AAM25448  
AC AAM25448 standard; protein; 153 AA.  
XX AAM25448;  
XX 16-OCT-2001 (first entry)  
DT Human protein sequence SEQ ID NO:963.  
DE  
KW Human; cancer; ulcer; HIV infection; human immunodeficiency virus;  
KW antiinflammatory; antirheumatic; antiarthritic; immunosuppressive;  
KW antibacterial; endocrine; cardiac; central nervous system; virucide;  
KW anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; anaemia;  
KW antiaggregant; haemostatic; vulnery; antilucer; osteopathic; eczema;  
KW dermatological; antiallergic; antiasthmatic; antidiabetic; cytostatic;  
KW neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;  
KW immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;  
KW antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;  
KW cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;  
KW genetic disease; haematopoietic disorder; platelet disorder; asthma;  
KW thrombocytopaenia; osteoporosis; severe combined immunodeficiency;  
KW allergic rhinitis; diabetes; multiple sclerosis; depression;  
KW Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;  
KW neurological disorder.  
XX  
OS Homo sapiens.  
XX  
XX WO200153455-A2.  
XX 26-JUL-2001.  
XX  
XX 22-DEC-2000; 2000WO-US035017.  
XX  
XX 23-DEC-1999; 99US-00471275.  
PR 21-JAN-2000; 2000US-00488725.  
PR 25-APR-2000; 2000US-00552317.  
XX  
XX (HYSE-) HYSEQ INC.  
XX  
XX Tang YT, Liu C, Drmanac RT;  
XX WPI; 2001-457603/49.  
DR N-PSDB; AAM99389.  
XX  
XX Isolated human polynucleotides encoding polypeptides, useful for the  
XX treatment and diagnosis of e.g. cancer, cancer, ulcers and HIV infection.  
XX  
XX Claim 20; Page 204; 1217pp; English.  
XX  
XX AAH99166 to AAH99904 encode the human proteins given in AAM25225 to  
XX AAM25963. The proteins can have activities based on the tissues and cells  
XX they are expressed in, such as: antinflammatory; antirheumatic;  
XX antiarthritic; immunosuppressive; antibacterial; endocrine; cardiac;  
XX central nervous system; virucide; anti-HIV; fungicide; antimutagen;  
XX cardiovascular; antianaemic; antiaggregant; haemostatic; vulnery;  
XX antilucer; osteopathic; dermatological; antiallergic; antiasthmatic;  
XX antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic;  
XX antiparkinsonian; and immunostimulant. The proteins and polynucleotides  
XX encoding them can be used in gene therapy, antisense therapy and vaccine  
XX production. The proteins and polynucleotides are useful for screening for  
XX agonists or antagonists of a protein and for the treatment and diagnosis  
XX of disorders associated with the activity of a protein e.g. inflammation,  
XX rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,  
XX neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal  
XX infections, autoimmunity, genetic diseases, haematopoietic disorders,

CC anaemia, platelet disorders, thrombocytopaenia, wounds, ulcers,  
CC osteoporosis, severe combined immunodeficiency, eczema, allergic  
CC rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,  
CC Alzheimer's disease, Parkinson's disease, neurodegenerative and  
CC neurological disorders  
XX  
XX Sequence 153 AA;  
Qy 548 DPYNILNFKAPFOPSGKNWKGRODSKASSEVTVPFSGHHTLLSCKAQLLESQDNCEGCG 607  
Db 4 DPYNILNFKAPFOPSGKNWKGRODSKASSEVTVPFSGHHTLLSCKAQLLESQDNCEGCG 63  
Qy 608 LGBALAGERVTHIKRKVTFLEBVTYYISGDDRGKWPWFARDGCRFQKRIQETVAI 667  
Db 64 QRDVLSGGRHTVKKVTFLEBVTYYISGDDRGKWPWFARDGCRFQKRIQETDAI 123  
Qy 668 GYCLAFEHREKMFNRLR 684  
Db 124 GYCLTFEHREMFNRLQ 140

RESULT 6  
ID ABB57373  
AC ABB57373 standard; protein; 657 AA.  
XX ABB57373;  
XX 07-MAR-2002 (first entry)  
DT  
XX  
XX Mouse ischaemic condition related protein sequence SEQ ID NO:1062.  
XX  
XX Mouse; ischaemia; compressive ischaemia; occlusive ischaemia;  
XX vasospastic ischaemia; ischaemic condition; ischaemic disease.  
XX  
XX Mus musculus.  
XX  
XX WO200188188-A2.  
XX  
XX 22-NOV-2001.  
XX  
XX 18-MAY-2001; 2001WO-JP004192.  
XX  
XX 18-MAY-2000; 2000JP-00145977.  
XX  
XX (YUNI-) UNIV NIHON SCHOOL JURIDICAL PERSON.  
XX  
XX Ishikawa K, Asai S, Takahashi Y, Nagata T, Ishii Y;  
XX WPI; 2002-034733/04.  
DR N-PSDB; ABI99909.  
XX  
XX Examining the ischemic condition (e.g. occlusive ischemia) by measuring  
XX expression levels of particular genes defined in the specification or by  
XX determining the expression profile of a gene group comprising these  
XX genes.  
XX  
XX Claim 2; Page 2676-2679; 2690pp; English.  
XX  
XX The present invention describes a method for examining ischaemic  
XX conditions, comprising measuring the expression levels of particular  
XX genes (I) in a test sample or determining the expression profile of a  
XX gene group in the sample comprising genes selected from (I). The method  
XX is useful for examining the ischaemic condition (e.g. compressive  
XX ischaemia, occlusive ischaemia or vasospastic ischaemia) by measuring  
XX expression levels of particular genes (ABI99202 to ABI99912, encoding the  
XX protein sequences in ABB57020 to ABB57374) or by determining the  
XX expression profile of a gene group comprising these genes. The expression  
XX levels or expression profiles produced by these genes are used as an  
XX indicator when screening for ischaemic condition-improving drugs or

CC therapeutics for ischaemic diseases. ABI99913 and ABI99914 represent PCR  
 CC primers for a mouse ischaemic condition related sequence, which are used  
 CC in the exemplification of the present invention  
 XX  
 SQ Sequence 657 AA;

Query Match 5.5%; Score 204; DB 5; Length 657;  
 Best Local Similarity 21.5%; Pred. No. 1.2e-08;  
 Matches 149; Conservative 86; Mismatches 290; Indels 168; Gaps 27;

QY 60 LLSQALLPSLFQKLLWSQSGGLIPTRWLDFAASYSALRASGRSESDAPTQVKSLS 119  
 DB 21 LLSPLMLLSR-----AWSRLRGPEVPEAWL-----AKVTGADQ 55  
 QY 120 YTAAGLFAKTRVYSTLALARG-----GTPVAVLVLRLEVKLKAQERAL---DSAAPTELL 171  
 DB 56 IEAALLTPFVSGNL-LPHGETESGSP-----EQSQAQRLCLVEAESPP----- 102  
 QY 172 EQQLMGVELLPS-SLQAGLVSHRELD-SSSGPLSVQSLG-----NFKVVS-LLN 219  
 DB 103 --ETWGLSNVDYNAKPGQDDLREKEMERTAGKATLQAPAGLQADKRLGEVVAEEGVAE 160  
 QY 220 PSYLDYLPQLGRLCQSSAGGQFVGFRYLTLPESCYSLEDGCHPQLRAEMSATWRRCPP 279  
 DB 161 PAY-----PTSQLEGGPAEN-----EEDGETVKTYQASAASIAPGYKPS 199  
 QY 280 LSTEGLEPEIHHRRMRWLFLQPNQODLPTLDQNGVHSLSEHNLLRMDPQHCTDNPAQ 339  
 DB 200 TPVPFLGAEHQATE-----EKGTEKADPNSSPGSHSRAWEY-YSREKPKQGEAKVE 254  
 QY 340 AVSPAADRP-----EPTTEKPELVIOEVSQSPQSSGLFCFELPVEKECEDHTNATDLSDRG 395  
 DB 255 AHRAGQGHPCRNAEAEEGGPETTP--VCTGNAFKAWVYRPGEDTEEDNSDSAEEDT 312  
 QY 396 ESLPVSTRPVCSNKLIDYILGAPSDLEASDSESED-----W-----GE 435  
 DB 313 AQTGATPHTSAFLKAWVYRPGEDTEEDSDSAEEDTAQTGATPHTSAFLKAWVYRPG 372  
 QY 436 EPEDDGFDSGLSSDVEQDSEGLHLWNSFHSVDPYKPNFTATIQTAARIAPRDPDS 495  
 DB 373 DTEENSDDL--AEEDTAQTGATPHT--SAFLKAWVYRPGEDTEEDNSDSAEEDTAQT 429  
 QY 496 GTWSGSCGVGSCQBPPLTPDHSSGDEEDWEPFSADEAKLKLWNSFCHSDPYNLNF 555  
 DB 430 GATPHTSPFLKAWVYRPGEDTEEDSENVAPGDSETADSSQSPC----- 478  
 QY 556 KAPFOPS-----GKNWKGQDSKASSEVTVFASGHH-----TLSCKAQLLES-- 598  
 DB 479 ---LQPRCLPGEKTKGRGEPPLFQVAFYLPGEKPSFWAAPKLPLRLQRLRLFKAPT 535  
 QY 599 --QEDNCPCGGLGEALAGERVTHIKKVKVTFLEEVTEYVIS-----GDEDRKGWEEFAR 651  
 DB 536 RDQDEIP-----LAKRVHAEKVTVHFLAWAGPAQAARRGWEEQAR 580  
 QY 652 DGCRCQRIQETEAIGYCLAPEHREKMFNRLR 684  
 DB 591 DRSRPARIAQAEEKLGPYLPDSSARAWARL 613

RESULT 7  
 ID AAY41102  
 XX AAY41102 standard; protein; 657 AA.  
 XX  
 AC AAY41102;  
 XX  
 DT 17-JAN-2000 (first entry)  
 XX  
 DE Rat MyD116 protein sequence.  
 XX  
 KW Progression-elevated gene-3; PEG-3 gene; regulatory region; cell death;  
 KW cell growth; cancer; thymidine kinase; gancyclovir; acyclovir; rat;  
 KW MyD116 protein.  
 XX

OS Rattus sp.  
 XX WO9949898-A1.  
 PN  
 XX 07-OCT-1999.  
 PD  
 XX 31-MAR-1999; 99WO-US0071199.  
 PF  
 XX 31-MAR-1998; 98US-00052753.  
 PR  
 XX (UYCO ) UNIV COLUMBIA NEW YORK.  
 PA  
 XX Fisher PB;  
 PI  
 XX WPI; 1999-591184/50.  
 DR  
 XX Novel vectors useful for studying the progression of cancer.  
 FT  
 XX Disclosure; Fig 3A-B; 251pp; English.  
 PS  
 XX The invention relates to an inducible progression-elevated gene-3 (PEG-3  
 CC gene) regulatory region functionally linked to a gene encoding a product  
 CC that causes or may be induced to cause the death or inhibition of cancer  
 CC cell growth. A vector of the invention which contains a gene encoding  
 CC thymidine kinase or a product which causes the cell to express a specific  
 CC antigen can be administered along with gancyclovir or acyclovir, or an  
 CC antibody or fragment to the antigen, respectively, to treat cancer in a  
 CC subject. The PEG-3 gene is useful for generating new cloning and  
 CC expression vectors, transfected cells, and for developing methods for  
 CC cultured growth of such cells. The PEG-3 polynucleotide is also useful as  
 CC a source of primers and probes to study the progression of cancer, and to  
 CC detect the presence of the gene. The present sequence represents the  
 CC amino acid sequence of the rat MyD116 protein  
 XX  
 SQ Sequence 657 AA;

Query Match 5.4%; Score 201; DB 2; Length 657;  
 Best Local Similarity 21.5%; Pred. No. 2.3e-08;  
 Matches 149; Conservative 85; Mismatches 291; Indels 168; Gaps 27;

QY 60 LLSQALLPSLFQKLLWSQSGGLIPTRWLDFAASYSALRASGRSESDAPTQVKSLS 119  
 DB 21 LLSPLMLLSR-----AWSRLRGPEVPEAWL-----AKVTGADQ 55  
 QY 120 YTAAGLFAKTRVYSTLALARG-----GTPVAVLVLRLEVKLKAQERAL---DSAAPTELL 171  
 DB 56 IEAALLTPFVSGNL-LPHGETESGSP-----EQSQAQRLCLVEAESPP----- 102  
 QY 172 EQQLMGVELLPS-SLQAGLVSHRELD-SSSGPLSVQSLG-----NFKVVS-LLN 219  
 DB 103 --ETWGLSNVDYNAKPGQDDLREKEMERTAGKATLQAPAGLQADKRLGEVVAEEGVAE 160  
 QY 220 PSYLDYLPQLGRLCQSSAGGQFVGFRYLTLPESCYSLEDGCHPQLRAEMSATWRRCPP 279  
 DB 161 PAY-----PTSQLEGGPAEN-----EEDGETVKTYQASAASIAPGYKPS 199  
 QY 280 LSTEGLEPEIHHRRMRWLFLQPNQODLPTLDQNGVHSLSEHNLLRMDPQHCTDNPAQ 339  
 DB 200 TPVPFLGAEHQATE-----EKGTEKADPNSSPGSHSRAWEY-YSREKPKQGEAKVE 254  
 QY 340 AVSPAADRP-----EPTTEKPELVIOEVSQSPQSSGLFCFELPVEKECEDHTNATDLSDRG 395  
 DB 255 AHRAGQGHPCRNAEAEEGGPETTP--VCTGNAFKAWVYRPGEDTEEDNSDSAEEDT 312  
 QY 396 ESLPVSTRPVCSNKLIDYILGAPSDLEASDSESED-----W-----GE 435  
 DB 313 AQTGATPHTSAFLKAWVYRPGEDTEEDSDSAEEDTAQTGATPHTSAFLKAWVYRPG 372  
 QY 436 EPEDDGFDSGLSSDVEQDSEGLHLWNSFHSVDPYKPNFTATIQTAARIAPRDPDS 495  
 DB 373 DTEENSDDL--AEEDTAQTGATPHT--SAFLKAWVYRPGEDTEEDNSDSAEEDTAQT 429  
 QY 496 GTWSGSCGVGSCQBPPLTPDHSSGDEEDWEPFSADEAKLKLWNSFCHSDPYNLNF 555

```

Db 430 GATPHTSPFLKAWYRFGEDTDDTEEDSENVAPOGSETADSSQSPC----- 478
Qy 556 KAPQPS-----GKWKGRQDSKASSEVTVAFSGHH-----TLLSCKAQLLES-- 598
Db 479 ---LQPORCLPGEYTKGRGEPPLFQVAFYLPGEKPSWNAAPKPLRLQRLRLFKAPT 535
Qy 599 --QEDNCPGCGGLGBALAGERYTHIKRKKVTFLEVTBYIIS-----GDEDRKGWEEFAR 651
Db 536 RDQDPEIP-----LKARKVHFAEKVTVHFLAVWAGPAQAARRGWEQFAR 580
Qy 652 DGCRFQRIQETVAIGCYCLAFEHREKMFNRLR 684
Db 581 DRSRFAARIAQAEKGLGYLTDPDSRARAWARLR 613

RESULT 8
AAW99891
ID AAW99891 standard; protein; 674 AA.
AC AAW99891;
DT 09-JUN-1999 (first entry)
DE Human growth regulator protein GRREG.
KW Human; growth regulator protein; GRREG; cancer.
OS Homo sapiens.
XX WO9902680-A1.
PN 21-JAN-1999.
PP 30-JUN-1998; 98WO-US013409.
PR 11-JUL-1997; 97US-00893852.
XX (INCY-) INCYTE PHARM INC.
XX Bandman O, Lal P, Shah P, Corley NC;
XX WPI; 1999-120881/10.
DR N-PSDB; AAX19798.
XX Novel human growth regulator protein - useful in the treatment of
XX cancers.
XX Claim 1; Fig 1; 70pp; English.
XX The present sequence represents a human growth regulator protein,
XX designated GRREG. The protein can be used in methods to treat cancer. The
XX protein can also be used in binding assays to detect (ant)agonists of
XX GRREG. A fragment from the nucleic acid sequence encoding GRREG can be
XX used as a probe for detecting GRREG encoding sequences (especially in PCR
XX amplified samples)
XX Sequence 674 AA;

Query Match 5.2%; Score 191.5; DB 2; Length 674;
Best Local Similarity 21.5%; Pred. No. 1.7e-07;
Matches 156; Conservative 75; Mismatches 268; Indels 227; Gaps 32;

Qy 60 LLSQALLPSLFQKLLWSQSLGJIPTR-WLDFPAASYSALRASRGREESDPTVQKSL 118
Db 21 LLSVMSLLSR-----AWSRLR-GLGLEPWLVAEVKGAAL----- 55
Qy 119 SYTAAGLFAKTRVVTALARGGTFVAVLRLVLEVKLKAQERALDSAAP-----T 168
Db 56 --VQAGLEGEAR-----TFLA--TPHTWGRPRPEEAEDSGGPGEDRETGLKT 100
Qy 169 FLLEQQWQVLLPSSLOAGLVSHRELDSSSSGSLPSVQSLGNFKVSYLLNPSYL----- 223

```

```

Db 101 SSSLPEAWG--LLDD--DDMGYGEREATSVPRQGSQFADQG-----RAPLSPSLLIRTQL 152
Qy 224 --DYLQGLRCCOSSAGGQGFVGRFLT-----PESCYLESDGCHPQPLRAEASATA--- 273
Db 153 GSDKNP--GEEKAEESGVAEEGVNKFSYPPSHRECCPAVEEEDDEEAVKKEAHTSTSA 210
Qy 274 -----WRRCP-----PLSTEGLEPEIHRRWRWLVFLQPNQOGDLPTLDQONGVHSL 320
Db 211 LSPGSKPSTWVSCPGSEENQATDKRTSKGARKTSVSPRSSGSDPRSWEYRSGEASEE 270
Qy 321 EENLLRMDPQHCTDNPQAVSPAADRPEPEKPELVIEQVSVSQSPQSGSLFCELPVEKE 380
Db 271 KE-----EKAHEETKGEAA-----PGQSAP-----AQRQLKSNWQCPSDEEE 311
Qy 381 CEEDHTNATDLSDRGESLPVSTRPVCNKLIDYILGAPSDLEASSDSESDWGEPEDD 440
Db 312 SEVKALGAAEKGEAEACPCPCIPPSAFLKAWVWPG--EDTEEEDEEE--DESDS 364
Qy 441 GFDSDGSLSESDVEQDSEGLHL--WNSFHSVDYKPNFTATIQ-----TAAKRAPRDP 493
Db 365 GSDEERGEAEASSSTPATGVFLKSW-----VYQFGEDTEEEDESDTGSADEREAE 417
Qy 494 DSCTSWSGCGVSCQEGPLPETPDH-----SSGEEDDWEPSADAENLKNWSPCHSE 547
Db 418 TSASTPPASAFKAWYRFGEDTEEEDEEDVDSEKDDSEALGAEAS----- 466
Qy 548 DYNLLNFKAPFQPSG-----KNWKGQDSKASSEVTVAFSGHHTLLSCKAQLLESQBDNC 603
Db 467 DPH-----PSHPDQSAHPRGHWYRPGKETEE-----EAEED-- 498
Qy 604 PGCLGEA-----LAGER-----YTHIKRK 623
Db 499 ---WGEAEPCPPRVAIVYVGEKPPPPWAPPRLRLQRLKRPETPTDHPDPTPLKAR 554
Qy 624 KYTFLEVTYIIS-----GDEDRKGPWEFADGCRFQKRIQETVAIGCYCLAFEHREK 678
Db 555 KVRPSEKVTVHFLAVWAGPAQAARQGWELQALDRSRFARRITQAOEELSPLCTPAAR 614
Qy 679 MFNRLR 684
Db 615 AWARLR 620

RESULT 9
AAV84366
ID AAV84366 standard; protein; 674 AA.
XX AAV84366;
AC AAV84366;
XX 12-JUL-2000 (first entry)
DT Amino acid sequence of a human apoptosis associated protein (GADD34).
DE Human; apoptosis associated protein; GADD34; antiviral agent;
XX growth arrest and DNA damage-inducible gene 34; viral infection;
KW serine/threonine phosphatase.
XX Homo sapiens.
XX GB2342716-A.
XX 19-APR-2000.
XX 27-JUL-1999; 99GB-00017631.
XX 14-SEP-1998; 98GB-00020025.
XX (PF1Z ) PFIZER LTD.
XX Ciaramella G;
XX WPI; 2000-285397/25.
DR N-PSDB; AAZ99766.

```





XX	SO	Sequence	674 AA;
Query Match			
Best Local Similarity 5.1%; Score 189.5; DB 7; Length 674;			
Matches 156; Conservative 74; Mismatches 269; Indels 227; Gaps 32;			
Qy	60	LLSQALLPSLFQKLLMSQSGGLIPTR-WLDFAAASYALRASRGRESDAPTQVQKSL	118
Db	21	LLSPVMGLLSR-----AWSRLR-GLGPLEPWLVEAVKGAAL-----	55
Qy	119	SYTAAGLFAKTRVVTLLALARGGTPVAVLRLVLEVKLKAQERALDSAAP-----T	168
Db	56	--VEAGLEGEAR-----TPLA--IPHTPWGRPEEAEADSGGCGEDRETGLKT	100
Qy	169	FLLEQQLWGVELLPSSLQAGLVSHRELDSSSGSLGKFNKVSYYLLNPSYL-----	223
Db	101	SSSLPEAWG--LLDD--DDGMYGREATSVPRGGSQFADGQ-----RAPLSPSLLIRTQ	152
Qy	224	--DYLPLGLRCQSSAGGQFVGFRILT-----PESCYLSEDCGCHPQPLRAEMSATA---	273
Db	153	GSDKNP--GEEKAEEGVAEEGVNKFSPSHRECCPAVEEEDDEAVKKEAHTTSTA	210
Qy	274	-----WRCRP-----PLSTEGLPFHRRMRWLFLQPNQGDLPDLDQNGYHSL	320
Db	211	LSPGKPEFTWVSCPEEENQATDKRTSKGAKTSPRSGSDPRSWEYRSGEASEE	270
Qy	321	BEHLLRMDPOHCTDNPAQAVSPAADRPETKPKPELVIEVVSQSGSLFCPELPVEKE	380
Db	271	KE-----EKAKETGGEAA-----PGQSSAP-----AORPQKSWQCPSEEE	311
Qy	381	CEEDHTNATLSDRGESLPVSTRVCSNKLIDYILGAPSDLEASSSESDGEEPEDD	440
Db	312	GEVKALGAERKDEAGCEPCIPPPSAFLKAWVWPG--EDTEEEDEEB-----DESDS	364
Qy	441	GFDSGSLSDVDQDSEGLHL--WNSFHSVDPKPQNFATIQ-----TAARIAPRDS	493
Db	365	GSDEEGEAEASSSTPATGVFLKSW-----VYQGEDTEEEDESDTGSAREDEAE	417
Qy	494	DSGTSWSGSCGVGSCQGLPETPDH-----SSGEEDDWEPSADAEENLKNWSPCHSE	547
Db	418	TSASTPPASAPLKAWVRPEDTEEEDEEDVDSEDKEDDEAALGEABS-----	466
Qy	548	DPYNLLNFKAPFPGS-----KNWGRQDSKASSEVTVAFGHHTLLSCKAQLLESQDNC	603
Db	467	DPH-----PSHDQRAHFRGWCYRPGKETEER-----EAEED--	498
Qy	604	PGCLGEA-----LAGER-----YTHIKRK	623
Db	499	----WGAEPCPFRAIVYVPEKEPPPPWAPPRLRLQRLKRPETPTHPDPETPLKAR	554
Qy	624	KVTFLEBVTYIIS-----GDEDRKGPWEFARDGCRFQKRIQTEVAIGVCLAFHREK	678
Db	555	KVRFSEKVTVHFLAVAGPAQAARQGPWEQLARDRSRFRRTITQAQEELSFCITPAAR	614
Qy	679	MFNRLR	684
Db	615	AWAELR	620
RESULT 11			
ADD45119 standard; protein; 674 AA.			
XX	AC	ADD45119;	
XX	XX	29-JAN-2004 (first entry)	
XX	XX	Human Protein XP_009097, SEQ ID NO 10552.	
XX	XX	Human; pain; neuronal tissue; gene therapy;	
XX	XX	spinal segmental nerve injury; chronic constriction injury; CCI;	
XX	XX	spared nerve injury; SNI; Chung.	

XX	OS	Homo sapiens.
XX	PN	WO2003016475-A2.
XX	XX	27-FEB-2003.
XX	XX	14-AUG-2002; 2002WO-US025765.
XX	XX	14-AUG-2001; 2001US-0312147P.
XX	XX	01-NOV-2001; 2001US-0346382P.
XX	XX	26-NOV-2001; 2001US-0333347P.
XX	XX	(GEHO ) GEN HOSPITAL CORP.
XX	XX	(FARB ) BAYER AG.
XX	PI	Woolf C, D'urso D, Befort K, Costigan M;
XX	XX	WPI; 2003-268312/26.
XX	XX	GENBANK; XP_009097.
XX	PT	New composition comprising two or more isolated polypeptides, useful for
XX	PT	preparing a medicament for treating pain in an animal.
XX	PS	Claim 1; Page; 1017pp; English.
XX	CC	The invention discloses a composition comprising two or more isolated rat
XX	CC	or human polynucleotides or a polynucleotide which represents a fragment,
XX	CC	derivative or allelic variation of the nucleic acid sequence. Also
XX	CC	claimed are a vector comprising the novel polynucleotide, a host cell
XX	CC	comprising the vector, a method for identifying a nucleotide sequence
XX	CC	which is differentially regulated in an animal subjected to pain and a
XX	CC	kit to perform the method, an array, a method for identifying an agent
XX	CC	that increases or decreases the expression of the polynucleotide sequence
XX	CC	that is differentially expressed in neuronal tissue of a first animal
XX	CC	subjected to pain, a method for identifying a compound which regulates
XX	CC	the expression of a polynucleotide sequence which is differentially
XX	CC	expressed in an animal subjected to pain, a method for identifying a
XX	CC	compound that regulates the activity of one or more of the
XX	CC	polynucleotides, a method for producing a pharmaceutical composition, a
XX	CC	method for identifying a compound or small molecule that regulates the
XX	CC	activity in an animal of one or more of the polypeptides given in the
XX	CC	specification, a method for identifying a compound useful in treating
XX	CC	pain and a pharmaceutical composition comprising the one or more
XX	CC	polypeptides or their antibodies. The polynucleotide or the compound that
XX	CC	modulates its activity is useful for preparing a medicament for treating
XX	CC	pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
XX	CC	injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. Gene
XX	CC	therapy). The sequence presented is a human protein (shown in Table 2 of
XX	CC	the specification) which is differentially expressed during pain. Note:
XX	CC	The sequence data for this patent did not form part of the printed
XX	CC	specification, but was obtained in electronic form directly from WIPO at
XX	CC	ftp.wipo.int/pub/published_pct_sequences.
XX	XX	Sequence 674 AA;
XX	XX	Query Match
XX	XX	Best Local Similarity 5.1%; Score 189.5; DB 7; Length 674;
XX	XX	Matches 156; Conservative 74; Mismatches 269; Indels 227; Gaps 32;
QY	60	LLSQALLPSLFQKLLMSQSGGLIPTR-WLDFAAASYALRASRGRESDAPTQVQKSL 118
DB	21	LLSPVMGLLSR-----AWSRLR-GLGPLEPWLVEAVKGAAL----- 55
QY	119	SYTAAGLFAKTRVVTLLALARGGTPVAVLRLVLEVKLKAQERALDSAAP-----T 168
DB	56	--VEAGLEGEAR-----TPLA--IPHTPWGRPEEAEADSGGCGEDRETGLKT 100
QY	169	FLLEQQLWGVELLPSSLQAGLVSHRELDSSSGSLGKFNKVSYYLLNPSYL----- 223
DB	101	SSSLPEAWG--LLDD--DDGMYGREATSVPRGGSQFADGQ-----RAPLSPSLLIRTQ 152
QY	224	--DYLPLGLRCQSSAGGQFVGFRILT-----PESCYLSEDCGCHPQPLRAEMSATA--- 273

```

Db 153 GSDKNP--GEKAEEGVAEEGVNKFSPYPPSHRECCPAVEEEDDEAVKKAHRTTSA 210
Qy 274 -----WRCPP-----PLSTEGLPETIHRRMRWLVPLOQNOQDLPTLDQNGYHSLE 320
Db 211 LSPGKPSFTWVSCFGEENQATDKRTSRGARKTSVSPRSSGSDPRSMWEYRSGEASEE 270
Qy 321 EEHLLRMDPOHCTDNPAAQVSPAARDPEPTEKKPELVIOEVSQSPQSGSLFCLPVEKE 380
Db 271 KE-----EKAKETGKEAA-----PGOSSAP-----AQRPLKSWWCQPSDEEE 311
Qy 381 CEEDHTNATDLSDRGESLPVSTRPVCSNKLIDYILGGAPSDLEASSDSESDWGEPEDD 440
Db 312 GEVKALGAAGKDEAGAECPPIPPPSAFLKAWYWP-----EDTEEEDEEE-----DESDS 364
Qy 441 GFDSGSLSESDVQDSEGLHL--WNSFHSVDPKPQNFTATIO-----TAARIAPRDP 493
Db 365 GSDEEGEAASSTPATGVFLKSW-----VYQGEDTEEEDESDTGSADEREAE 417
Qy 494 DSGTSWSGCGVSCQEGPLPETPDH-----SSGEEDDWEPSADEAENLKLWNSFCHSE 547
Db 418 TSASTPPASAPLKAWYRPGEDTEEEEDVDSEKEDDSEALGEAES-----466
Qy 548 DPYNLLNFKAPFQSG-----KNWGRQDSKASSEVTVAFSGHHTLLSCKAQLLESQEDNC 603
Db 467 DPH-----PSHPDQRAHFRGWYRPGKETEEE-----EAEED--498
Qy 604 PCCGLGEA-----LAGER-----YTHIKRK 623
Db 499 -----WGEABCPFRVAYVPGKEPPPWAPRLPLRLQRLKRPTETPHDPDPETPLKAR 554
Qy 624 KVTLEEVTEYVIS-----GBEDRKGPMEEFARDGCRFOKRIQETEAIVAGYCLAFEHREK 678
Db 555 KVRSEKTVHFLAVWAGPAQAARQGPWEQLARDSRFARITQAEBSLCLTPAAR 614
Qy 679 MNRLR 684
Db 615 AWARLR 620

RESULT 12
ID ADE54744 standard; protein; 674 AA.
XX AC ADE54744;
XX DT 29-JAN-2004 (first entry)
XX DE Human Protein XP_009097, SEQ ID NO 549.
XX KW Human; pain; neuronal tissue; gene therapy;
XX KW spinal segmental nerve injury; chronic constriction injury; CCI;
XX KW spared nerve injury; SNI; Chung.
XX OS Homo sapiens.
XX PN WO2003016475-A2.
XX PD 27-FEB-2003.
XX PF 14-AUG-2002; 2002WO-025765.
XX PR 14-AUG-2001; 2001US-0312147P.
XX PR 01-NOV-2001; 2001US-0346382P.
XX PR 26-NOV-2001; 2001US-0333347P.
XX PA (GENO ) GEN HOSPITAL CORP.
XX PA (FARB ) BAYER AG.
XX PI Woolf C, D'urso D, Befort K, Costigan M;
XX WPI; 2003-268312/26.
XX DR GENBANK; XP_009097.

```

```

XX New composition comprising two or more isolated polypeptides, useful for
Pt preparing a medicament for treating pain in an animal.
XX Claim 1; Page; 1017pp; English.
XX The invention discloses a composition comprising two or more isolated rat
Cc or human polynucleotides or a polynucleotide which represents a fragment,
Cc derivative or allelic variation of the nucleic acid sequence. Also
Cc claimed are a vector comprising the novel polynucleotide, a host cell
Cc comprising the vector, a method for identifying a nucleotide sequence
Cc which is differentially regulated in an animal subjected to pain and a
Cc kit to perform the method, an array, a method for identifying an agent
Cc that increases or decreases the expression of the polynucleotide sequence
Cc that is differentially expressed in neuronal tissue of a first animal
Cc subjected to pain, a method for identifying a compound which regulates
Cc the expression of a polynucleotide sequence which is differentially
Cc expressed in an animal subjected to pain, a method for identifying a
Cc compound that regulates the activity of one or more of the
Cc polynucleotides, a method for producing a pharmaceutical composition, a
Cc method for identifying a compound or small molecule that regulates the
Cc activity in an animal of one or more of the polypeptides given in the
Cc specification, a method for identifying a compound useful in treating
Cc pain and a pharmaceutical composition comprising the one or more
Cc polypeptides or their antibodies. The polynucleotide or the compound that
Cc modulates its activity is useful for preparing a medicament for treating
Cc pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
Cc injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
Cc therapy). The sequence presented is a human protein (shown in Table 2 of
Cc the specification) which is differentially expressed during pain. Note:
Cc The sequence data for this patent did not form part of the printed
Cc specification, but was obtained in electronic form directly from WIPO at
Cc ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 674 AA;

Query Match 5.1%; Score 189.5; DB 7; Length 674;
Best Local Similarity 21.5%; Pred. No. 2.6e-07;
Matches 156; Conservative 74; Mismatches 269; Indels 227; Gaps 32;

Qy 60 LLSQLLALLPSLQKLLNLSQSLGGLIPTR-WLDFAAYSALRASRGRESDAPTQKSL 118
Db 21 LLSFVPMGLLSR-----AWSRLR-GLGFLEPWLVEAVKGAAL-----55
Qy 119 SYTAAGLFAKTRVVTTLALARGGTPVAVLVLRLEVKLKAQALDLSAAP-----T 168
Db 56 --VEAGLEGEAR-----TPLA--IPHTPWGRPEEAEADSGGREDRETLGKT 100
Qy 169 FLLEQQLWGVLLPSSLQAGLVSHRELDSSSGPLSVQSLGNFKVSYLYLNPSYL-----223
Db 101 SSSLPEAWG--LLDD--DDGMGEREATSVPRGQSGQFADGQ----RAPLSPLLIRTLQ 152
Qy 224 --DYLPLQLGRCCSSAGGQFVGRITL-----PESCYLSEDCGHPQLAENSATA---273
Db 153 GSDKNP--GEKAEEGVAEEGVNKFSPYPPSHRECCPAVEEEDDEAVKKAHRTTSA 210
Qy 274 -----WRCPP-----PLSTEGLPETIHRRMRWLVPLOQNOQDLPTLDQNGYHSLE 320
Db 211 LSPGKPSFTWVSCFGEENQATDKRTSRGARKTSVSPRSSGSDPRSMWEYRSGEASEE 270
Qy 321 EEHLLRMDPOHCTDNPAAQVSPAARDPEPTEKKPELVIOEVSQSPQSGSLFCLPVEKE 380
Db 271 KE-----EKAKETGKEAA-----PGOSSAP-----AQRPLKSWWCQPSDEEE 311
Qy 381 CEEDHTNATDLSDRGESLPVSTRPVCSNKLIDYILGGAPSDLEASSDSESDWGEPEDD 440
Db 312 GEVKALGAAGKDEAGAECPPIPPPSAFLKAWYWP-----EDTEEEDEEE-----DESDS 364
Qy 441 GFDSGSLSESDVQDSEGLHL--WNSFHSVDPKPQNFTATIO-----TAARIAPRDP 493
Db 365 GSDEEGEAASSTPATGVFLKSW-----VYQGEDTEEEDESDTGSADEREAE 417
Qy 494 DSGTSWSGCGVSCQEGPLPETPDH-----SSGEEDDWEPSADEAENLKLWNSFCHSE 547

```

Db 418 TSASTPPASAFKAWVYRPGEDTDEEDDDSEALGEAES----- 466  
Qy 548 DPNYLLNFKAPFQPSG----KNWKGROQSKASSVTVAFSGHHTLLSCKAQLLESQDNDC 603  
Db 467 DPH-----PSHPDQRAHFGWYRPGKETEER-----EAEED-- 498  
Qy 604 PGCSLGEA-----LAGER-----YTHIKRK 623  
Db 499 ----WGEAEPCEPFRVAIYVPGKPPPPWAPRLRLQRLKRPETPTDPTPLKAR 554  
Qy 624 KVTFLBETVEYIS-----GDEDRKGPWEEFARDGCRFKRIQIETVAIGYCLAFEHREK 678  
Db 555 KVRSEKVTYVHFLAVWAGPAQAARQGWEOQLARDRSFARRITOAQBELSPCLTPAAR 614  
Qy 679 MFNRLR 684  
Db 615 AWARLR 620

RESULT 13  
AAB92888  
ID AAB92888 standard; protein; 674 AA.  
XX  
AC AAB92888;  
XX  
DT 26-JUN-2001 (first entry)  
XX  
DE Human protein sequence SEQ ID NO:11494.  
XX  
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.  
XX  
OS Homo sapiens.  
XX  
PN EP1074617-A2.  
XX  
PD 07-FEB-2001.  
XX  
PF 28-JUL-2000; 2000EP-00116126.  
XX  
PR 29-JUL-1999; 99JP-00248036.  
XX  
PR 27-AUG-1999; 99JP-00300253.  
XX  
PR 11-JAN-2000; 2000JP-00118776.  
XX  
PR 02-MAY-2000; 2000JP-00183767.  
XX  
PR 09-JUN-2000; 2000JP-00241899.  
XX  
PA (HELI-) HELIX RES INST.  
XX  
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
PI Iehii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX  
XX WPI; 2001-318749/34.  
XX  
PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.  
XX  
PS Claim 8; SEQ ID NO 11494; 2537pp + Sequence Listing; English.  
XX  
CC The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in

CC gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAB92888 and AAB92889 represent human amino acid sequences; AAB92890 to AAB92893 represent human amino acid sequences; and AAB92894 to AAB92897 represent present invention  
XX  
SQ Sequence 674 AA;  
Query Match 5.0%; Score 186.5; DB 4; Length 674;  
Best Local Similarity 21.5%; Pred. No. 4.9e-07;  
Matches 156; Conservative 73; Mismatches 270; Indels 227; Gaps 32;  
Qy 60 LLSOLLALLPSLFQKLLMSQLSGGLIPTR-WLDEAASYSALRASRGRESDAPTVQKSL 118  
Db 21 LLSFVWGLLSR-----AWSRLR-GLGPLEPWLVEAVKGAAL----- 55  
Qy 119 SYTAAGLFAKTRVYSTLALARGGTPVAVLRLVLEVKLKAQERALDSAAP-----T 168  
Db 56 --VEAGLEGEAR-----TPLA--IPHTPWGRRPGCEAEADSGGPGEDRETGLKLT 100  
Qy 169 FLLEQQLWGVLLPSSLOAGLVSHRELDSSSSSPLSVQSLGNFKVSYLLNPSYL----- 223  
Db 101 SSSLPEAWG--LLDD--DDGMYGEREATSVPRGQSGQFADGQ----RAPLSPSLIRTQL 152  
Qy 224 --DYLQGLRCQSSAGGQFVGFRTLT-----PESCYLSEDCGCHPOPLRAENSATA-- 273  
Db 153 GSDKNP--GEEKAELEGVAEEGVNKFSPYPSHRECCPAVEEEDDEAVKEAHRSTSA 210  
Qy 274 -----WRRCP-----PLSTEGPLPEIHHRRMRVLVFLQPNQOQDLPTLDOONGYHSL 320  
Db 211 LSPGSKPSTWVSCPGEEENQATDKRTSKGARKTSVSPRSSGSDPSRWEYSRGEASEE 270  
Qy 321 EENLLRMDPOHCTDNPAQAVSPAARDPEPTEKPELVLTQEVVSQSGSLFELPELVEKE 380  
Db 271 KE-----EKAHEETGKEAA-----PGQSSAP-----AQRPLKSWWCQPSDEE 311  
Qy 381 CBEDHTNATDLSDRGESLPVSTRPVCSNKLDIYILGAPSDLEASDSESDNGERPEDD 440  
Db 312 SEVKPLGAAEKOGAEACPPCIPPPSAFLKAWVTWPG---EDTEEEDEEE---DESDS 364  
Qy 441 GFDSGSLSESDEVEQDSGLHL--MNSFHSVDPYKPNFTATIQ-----TAARIAPRDPS 493  
Db 365 GSDEESGEAEASSSTPATGVFLKSW-----VYQGEDTEEBEEDSDTGSADEREAE 417  
Qy 494 DSGTWSGCGVSGCGEGPLPETPDH-----SSGEEDWEPSEADAEENLKUNSPCHSE 547  
Db 418 TSASTPPASAFKAWVYRPGEDTEEBEEDDDSEALGEAES----- 466  
Qy 548 DPNYLLNFKAPFQPSG----KNWKGROQSKASSVTVAFSGHHTLLSCKAQLLESQDNDC 603  
Db 467 DPH-----PSHPDQRAHFGWYRPGKETEER-----EAEED-- 498  
Qy 604 PGCSLGEA-----LAGER-----YTHIKRK 623  
Db 499 ----WGEAEPCEPFRVAIYVPGKPPPPWAPRLRLQRLKRPETPTDPTPLKAR 554  
Qy 624 KVTFLBETVEYIS-----GDEDRKGPWEEFARDGCRFKRIQIETVAIGYCLAFEHREK 678  
Db 555 KVRSEKVTYVHFLAVWAGPAQAARQGWEOQLARDRSFARRITOAQBELSPCLTPAAR 614  
Qy 679 MFNRLR 684  
Db 615 AWARLR 620

RESULT 14  
ADRI4089  
ID ADRI4089 standard; protein; 674 AA.  
XX



```

245 Db EGETKPBQHRAGQSHP-----CQNAESE---EGGPETSCSGSAFLKAWVYRGEDETEE 295
Qy 440 DGFDSDSGLSESDVEQSEGLHLWNSPHSVDPYKPPQNFTATQTAARAPRDPDSDSGTSM 499
Db 296 EE-DPDLDSABEDTAHTCTTPT-SAFLKAWVYRPGEDT-----EEDDGG-DW 340
Qy 500 SGSCGVCSCQEGPLPET-----PDHSSGEEDDWEPSAD-EAENKLWNSFCH--- 545
Db 341 D-SABEDAAQSCCTTPHTSAFLKAWVYRPGEDTEEDDSENPAVDSTVSCQCTQHCLP 399
Qy 546 -----SEDYPNLLNFKAPQFQSPGKNWKGRODSKASSEVTVAFSGHITLLSCKAQLL 596
Db 400 VEKTKGGEAEPPFPQVAFYLPQKPPW-----AAPKLPLRL--QKELRSFKA--- 447
Qy 597 ESQEDNCPGGLGEALAGERYTHIKRKYTFLEBEVEYYIS-----GDEDRKGWSEFAR 651
Db 448 -----PARNQGPPIP-----LKGKRVHFSEKTVHFLAVWAGPAQAARRGPEQFAR 494
Qy 652 DGRFQKRIQIETEVAIQCYCLAFEHREKWFNRLR 684
Db 495 DRSRFARRIAQAEQLGPYLTPAFRARAWTLR 527

Search completed: September 16, 2005, 10:42:36
Job time : 74.7612 secs

```

KW	DNA repiar; tumorigenesis; angiogenesis; metastasis; melanoma;
KW	brain cancer; cervix cancer; prostate cancer; lung cancer;
KW	colorectal cancer; neuroblastoma; glioblastoma; diagnosis; therapy.
XX	
OS	Homo sapiens.
XX	
XX	WO9842315-A1.
PN	
XX	
PD	01-OCT-1998.
XX	
PF	20-MAR-1998; 98WO-US005793.
XX	
PR	21-MAR-1997; 97US-00821818.
XX	
PA	(UYCO ) UNIV COLUMBIA NEW YORK.
XX	
PA	(GENQ-) GENQUEST INC.
PI	Fisher PB, Goldstein NI, Su Z, Zhang N;
DR	
XX	WPI; 1998-557025/47.
XX	N-PSDB; AAV65765.
XX	
PT	New isolated Progression Elevated Gene-3 - used to develop products for
PT	e.g. modulating DNA damage and repair pathways, cancer progression or
PT	oncogene mediated transformation and angiogenesis.
XX	
PS	Claim 17; Fig 13A-C; 235pp; English.
XX	
CC	This polypeptide comprises human progression elevated gene-3 (PEG-3)
CC	protein. The amino acid sequence was deduced from an isolated cDNA clone
CC	(see AAV65765). PEG-3 expression (1.5 and 2.8 kb transcripts) is elevated
CC	in most human tumour cell lines. The amount of PEG-3 in a cell is also an
CC	indicator of DNA damage in that cell. Cells in which a reporter gene in
CC	under control of the promoter or regulatory region (see AAV65766) of the
CC	rat PEG-3 gene can be used in methods for identifying agents that
CC	modulate PEG-3 expression or the ability of PEG-3 to induce progression,
CC	or for determining whether an agent is capable of inhibiting DNA damage,
CC	and repair pathways, cancer progression or oncogene-mediated
CC	transformation. Compounds that induce DNA damage or which regulate
CC	angiogenesis can also be identified using such cells. Transgenic animals
CC	and vaccines comprising PEG-3 polypeptides and an immune response
CC	enhancer are also claimed. Antibodies (especially monoclonal) to the PEG-
CC	3 may be used to determine tumour progression of a cell. Cells can be
CC	protected from chemotherapeutic damage by inhibiting or eliminating the
CC	expression of PEG-3 in the cells. The methods can be applied to a
CC	progression phenotype comprising anchorage-independent growth,
CC	tumorigenesis, angiogenesis or metastasis, to melanoma, brain, cervical,
CC	prostate, lung or colorectal cancer, neuroblastoma or glioblastoma (all
CC	claimed)
XX	
SQ	Sequence 578 AA;
	Query Match 4.9%; Score 182.5; DB 2; Length 578;
	Best Local Similarity 22.7%; Pred. No. 8.7e-07;
	Matches 130; Conservative 61; Mismatches 213; Indels 169; Gaps 28
Qy	216 YLLNPSYLDLPQLGLR-----CQSAGGGQFVGRTRLTPESCYSLEDGCHPOP 264     : : :   : :     : : :   : :   : :   : :
Db	20 YLLSP-LMGFLRSRAWRLRGPEVSEAWLAETVAGANQTADALLTPPP--VSENHL---P 73     : : :   : :     : : :   : :   : :   : :
Qy	265 LR-----AESMTAWRRRC-----PPIS-----TEGLPEIHRRMRW 295     : : :   : :     : : :   : :   : :   : :
Db	74 LRETEGNTPESWAKAAQLCLDVEAQSPPKTKMGLSDIDHNKPKQDGREQVEHTAG 133     : : :   : :     : : :   : :   : :   : :
Qy	296 LVFLQP--NQGD-----LPTLDQNGYHSLREHNLRLMDPOHCITDN 336             : : :     : : :   : :   : :   : :
Db	134 LPTLQPLHQADKKVGSWAREEGVSELAYPTSHWEGPADDEDTFTVKKAHQASAS 193             : : :     : : :   : :   : :   : :
Qy	337 PAQAVSPA-----ADRPETEKPELVIQEVSQSQGS-SLFCELPLVSKCEE 383     : : :   : :     : : :   : :   : :   : :
Db	194 IAPGYKSTSVYCGEAHRATEEKGTONKAE-----PSGHSRFWEYHTERPKQ 244     : : :   : :     : : :   : :   : :   : :
Qy	384 DHTNATDLSDRGESLPVSTRPVCSNKLI DYIIGGAPS DLEASSDSESDDW----GEEPED 439     : : :   : :     : : :   : :   : :   : :

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 16, 2005, 09:17:12 ; Search time 18.798 Seconds  
(without alignments)  
2771.838 Million cell updates/sec

Title: US-10-650-482-4

Perfect score: 3710

Sequence: 1 METGTHRRARXPGPRLGSWF.....MPNRLRIESKDLLLYSNVKK 698

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : Issued Patents AA.\*

1: /cgn2\_6/ptodata/1/1aa/5A COMB.pep.\*

2: /cgn2\_6/ptodata/1/1aa/5B COMB.pep.\*

3: /cgn2\_6/ptodata/1/1aa/6A COMB.pep.\*

4: /cgn2\_6/ptodata/1/1aa/6B COMB.pep.\*

5: /cgn2\_6/ptodata/1/1aa/PCTUS COMB.pep.\*

6: /cgn2\_6/ptodata/1/1aa/backfiles.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	204	5.5	657	3	US-08-893-852A-3
2	204	5.5	657	3	US-08-821-818-3
3	204	5.5	657	4	US-09-052-753B-3
4	191.5	5.2	674	3	US-08-893-852A-1
5	182.5	4.9	578	4	US-09-052-753B-7
6	171.5	4.6	590	3	US-08-893-852A-4
7	171.5	4.6	590	3	US-08-821-818-2
8	171.5	4.6	590	4	US-09-052-753B-2
9	132.5	3.6	1065	4	US-09-949-016-11618
10	127.5	3.4	3913	4	US-09-949-016-10933
11	127	3.4	1503	3	US-08-976-255-14
12	125	3.4	2781	4	US-09-698-295-10
13	125	3.4	2907	4	US-09-698-295-1
14	124.5	3.4	1780	3	US-08-769-309A-5
15	124.5	3.4	1780	3	US-08-994-570-5
16	123.5	3.3	4377	4	US-09-949-016-6978
17	123	3.3	1142	2	US-08-993-118-7
18	123	3.3	1142	3	US-08-845-528C-7
19	123	3.3	1142	3	US-09-061-709-2
20	123	3.3	1142	3	US-09-066-281B-7
21	123	3.3	1142	4	US-09-899-651-2
22	123	3.3	1142	4	US-09-468-433C-7
23	123	3.3	1142	4	US-09-392-714-26
24	123	3.3	1142	4	US-09-270-437D-2
25	120.5	3.2	1133	4	US-09-252-991A-17945
26	120	3.2	1261	4	US-09-949-016-9651
27	119.5	3.2	961	4	US-09-538-092-1231

28	119	3.2	414	4	US-09-248-796A-19046	Sequence 19046, A
29	118	3.2	1781	4	US-09-961-403-13	Sequence 13, Appl
30	117.5	3.2	1712	4	US-09-949-016-9450	Sequence 9450, Ap
31	117	3.2	1234	5	PCT-US95-13041-15	Sequence 15, Appl
32	117	3.2	1234	5	PCT-US95-13041-15	Sequence 15, Appl
33	117	3.2	1317	3	US-09-083-521-7	Sequence 7, Appl
34	116.5	3.1	912	5	PCT-US95-03747-2	Sequence 2, Appl
35	115	3.1	805	4	US-09-344-624-19	Sequence 19, Appl
36	115	3.1	805	4	US-09-949-016-6713	Sequence 6713, Ap
37	115	3.1	3340	4	US-09-252-991A-23568	Sequence 23568, A
38	114.5	3.1	654	1	US-08-083-590A-16	Sequence 16, Appl
39	114.5	3.1	654	2	US-08-346-128-35	Sequence 35, Appl
40	114.5	3.1	654	3	US-08-532-384-16	Sequence 16, Appl
41	114.5	3.1	1187	1	US-08-320-559-28	Sequence 28, Appl
42	114.5	3.1	1187	3	US-08-545-860D-28	Sequence 28, Appl
43	114.5	3.1	1187	5	PCT-US94-04496-28	Sequence 28, Appl
44	114.5	3.1	1210	1	US-08-320-559-26	Sequence 26, Appl
45	114.5	3.1	1210	3	US-08-545-860D-26	Sequence 26, Appl
46	114.5	3.1	1210	5	PCT-US94-04496-26	Sequence 1179, Ap
47	114.5	3.1	1210	5	PCT-US94-04496-26	Sequence 26, Appl
48	114.5	3.1	2523	1	US-08-185-432-18	Sequence 18, Appl
49	114.5	3.1	2523	4	US-08-899-232-3	Sequence 3, Appl
50	114.5	3.1	2523	4	US-09-121-457-3	Sequence 3, Appl
51	114	3.1	605	3	US-09-394-645-2	Sequence 2, Appl
52	114	3.1	605	3	US-09-243-560B-2	Sequence 2, Appl
53	114	3.1	848	3	US-08-976-255-10	Sequence 10, Appl
54	114	3.1	1219	4	US-09-344-624-4	Sequence 4, Appl
55	113.5	3.1	445	3	US-08-845-258-38	Sequence 38, Appl
56	113.5	3.1	445	3	US-08-990-571-38	Sequence 38, Appl
57	113.5	3.1	445	3	US-08-723-142A-38	Sequence 38, Appl
58	113.5	3.1	445	4	US-09-528-784A-38	Sequence 38, Appl
59	113.5	3.1	445	4	US-09-569-098A-38	Sequence 38, Appl
60	113.5	3.1	605	4	US-09-949-016-11347	Sequence 11347, A
61	113.5	3.1	666	4	US-09-528-784A-85	Sequence 85, Appl
62	113.5	3.1	666	4	US-09-569-098A-85	Sequence 85, Appl
63	113.5	3.1	1132	4	US-09-528-784A-87	Sequence 87, Appl
64	113.5	3.1	1132	4	US-09-569-098A-87	Sequence 87, Appl
65	112.5	3.0	1076	4	US-09-949-016-7421	Sequence 7421, Ap
66	112.5	3.0	1805	1	US-07-853-913-2	Sequence 2, Appl
67	112	3.0	576	2	US-09-367-206-22	Sequence 22, Appl
68	112	3.0	1183	2	US-08-447-031A-2	Sequence 2, Appl
69	112	3.0	2353	3	US-08-984-709A-50	Sequence 50, Appl
70	111.5	3.0	377	4	US-09-248-796A-20227	Sequence 20227, A
71	111.5	3.0	619	3	US-09-066-046-2	Sequence 2, Appl
72	111.5	3.0	773	4	US-09-270-767-33141	Sequence 33141, A
73	111.5	3.0	773	4	US-09-270-767-48358	Sequence 48358, A
74	111.5	3.0	1969	4	US-09-418-710-72	Sequence 72, Appl
75	111.5	3.0	1969	4	US-09-839-479-71	Sequence 71, Appl
76	111	3.0	576	3	US-09-367-206-1	Sequence 1, Appl
77	111	3.0	576	3	US-09-367-206-21	Sequence 21, Appl
78	111	3.0	576	3	US-09-367-206-23	Sequence 23, Appl
79	111	3.0	581	4	US-09-244-805-27	Sequence 27, Appl
80	111	3.0	1155	1	US-08-094-948A-29	Sequence 29, Appl
81	111	3.0	1155	5	PCT-US96-03319-29	Sequence 29, Appl
82	110.5	3.0	2870	4	US-09-479-467A-15	Sequence 15, Appl
83	110.5	3.0	3178	4	US-09-479-467A-4	Sequence 4, Appl
84	110	3.0	484	3	US-09-066-046-8	Sequence 8, Appl
85	110	3.0	489	1	US-07-903-103-4	Sequence 4, Appl
86	110	3.0	489	1	US-08-044-619A-4	Sequence 4, Appl
87	110	3.0	489	1	US-08-283-911-4	Sequence 4, Appl
88	110	3.0	489	1	US-08-245-500A-5	Sequence 5, Appl
89	110	3.0	489	1	US-08-390-546-5	Sequence 5, Appl
90	110	3.0	489	1	US-08-390-479A-5	Sequence 5, Appl
91	110	3.0	489	1	US-08-557-393-5	Sequence 5, Appl
92	110	3.0	489	1	US-08-350-516C-5	Sequence 5, Appl
93	110	3.0	489	1	US-08-350-517A-5	Sequence 5, Appl
94	110	3.0	489	1	US-08-801-718-5	Sequence 5, Appl
95	110	3.0	489	3	US-09-170-159A-5	Sequence 5, Appl
96	110	3.0	489	3	US-09-480-718-46	Sequence 46, Appl
97	110	3.0	544	3	US-09-087-134-14	Sequence 14, Appl
98	110	3.0	544	3	US-09-949-016-6963	Sequence 6963, Ap
99	110	3.0	757	4	US-09-949-016-8087	Sequence 8087, Ap
100	110	3.0	758	4	US-09-949-016-8087	Sequence 8087, Ap

## ALIGNMENTS

## RESULT 1

US-08-893-852A-3  
 ; Sequence 3, Application US/08893852A  
 ; Patent No. 6080558  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Bandman, Olga  
 ; APPLICANT: Lal, Preeti  
 ; APPLICANT: Shah, Purvi  
 ; APPLICANT: Corley, Neil C.  
 ; TITLE OF INVENTION: NEW HUMAN GROWTH REGULATOR PROTEIN  
 ; NUMBER OF SEQUENCES: 4  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
 ; STREET: 3174 Porter Drive  
 ; CITY: Palo Alto  
 ; STATE: CA  
 ; COUNTRY: USA  
 ; ZIP: 94304  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette  
 ; COMPUTER: IBM Compatible  
 ; OPERATING SYSTEM: DOS  
 ; SOFTWARE: Fast-Seq for Windows Version 2.0  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/893,852A  
 ; FILING DATE: Filed Herewith  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER:  
 ; FILING DATE:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Billings, Lucy J.  
 ; REGISTRATION NUMBER: 36,749  
 ; REFERENCE/DOCKET NUMBER: PF-0341 US  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 415-855-0555  
 ; TELEFAX: 415-845-4166  
 ; INFORMATION FOR SEQ ID NO: 3:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 657 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; IMMEDIATE SOURCE:  
 ; LIBRARY: GenBank  
 ; CLONE: 53041  
 ; US-08-893-852A-3

Query Match 5.5%; Score 204; DB 3; Length 657;

Best Local Similarity 21.5%; Pred. No. 9.4e-10;  
 Matches 149; Conservative 86; Mismatches 290; Indels 168; Gaps 27;

QY 60 LLSQLALLPSLFOKLLWSQSGGLIPTRWLDPAASYSALRASGRRESDAPTVOQSLS 119  
 DB 21 LLSPLMGLLSR-----NWSRLRGPEVPEAWL-----AKTVTGADQ 55  
 QY 120 YTAAGLFAKTRVSTLALARG-----GTPVAVLVRLEVLKKAQERAL---DSAAPTFL 171  
 DB 56 IEMALLTPTVSGNL-LPHGETESGSPF-----EQSQAQRCLCLVEAESPP----- 102  
 QY 172 EQQLWGVLLPS-SIQAGLVSHRELD-SSSSGPLSVQSLG-----NFKVSVY---LLN 219  
 DB 103 --ETWGLSNVDEYNAPKQODLREKEMERTAGKATLPAGLQAGADKRLGEVVAEEGVAE 160  
 QY 220 PSYLDLPQLGLRCQSSAGGQFVGFRLLTPESCYSLSGDCGHPOPLRAEMSATAWRCPP 279  
 DB 161 PAY-----PTSQLEGGPAEN-----EEDGETVTKYQASASIAPIGYKPS 199  
 QY 280 LSTGLPEIHHRRMRWLVLQPNQGDLPITLDQONGYHLSSEHNLLRMDPQHCTDNPQ 339

DB 200 TPVPFLGEAEHQATE-----EKGTEKADPSNPSGSGSHRAWY-YSRKEPKQGEAKVE 254  
 QY 340 AVSPAADRP-----PTEKKKPELVIQEYSQSPQSSLFCELPVEKECEDHTNATDLSDRG 395  
 DB 255 AHRAGQGHPCRNABAEAGPETTF-VCTGNAFKAWVYRPGEDTEEDSDSDSAEEDT 312  
 QY 396 ESLPVSTRPVCSNKLIDYILGAPSDLEASDSESD-----N-----GE 435  
 DB 313 AQTGATHTSAFLKAWYRPGEDTEEDSDSDSAEEDTAQTGATPHTSAFLKAWYRPG 372  
 QY 436 EPEDGFDSDGSLSESDVQDSEGLHLWNSFHSVDPYKQNFATTIOTAARIAPRPSDS 495  
 DB 373 DTEENSDDLDS--AEEDTAQTGATPHT-SAPLKAWVYRPGEDTEEDSDSDSAEEDTAQT 429  
 QY 496 GTSWSGCGVSGCOEGLPETPDHSSGEEDDWEPSADEAENKLWNSFCHSEDYNNLNF 555  
 DB 430 GATPHTSPFLKAWVYRPGEDTDEEDSDSENVAPGDSADSSQSPC----- 478  
 QY 556 KAPFQPS-----GKNWKGRODSKASSEVTVAFSGHH-----TLLSCKAQLLES-- 598  
 DB 479 ---LQPORCLPGEKTKGRGEPPLFQVAFYLPGEKPSWAAAPKLPLRLQRLRLFKAPT 535  
 QY 599 --QEDNCPGCGGLGALAGERYTHIKRKYVTFLEEVTEYIIS-----GDEDRKGWESFAR 651  
 DB 536 RDQDPEIP-----LKARKVHFAEKVTVHFLAVWAGPAQAARRGPEWQFAR 580  
 QY 652 DGCRRFQRIQETVEAIGCYCLAFEHREKMFNRLR 684  
 DB 581 DRSRFAIRIAQAEEKLGLPYLTPOSRARAWALR 613

## RESULT 2

US-08-821-818-3  
 ; Sequence 3, Application US/08821818  
 ; Patent No. 6146877  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Fisher, Paul B.  
 ; TITLE OF INVENTION: IDENTIFICATION OF THE PROGRESSION  
 ; TITLE OF INVENTION: ELEVATED GENE-3 AND USES THEREOF  
 ; NUMBER OF SEQUENCES: 5  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Cooper & Dunham LLP  
 ; STREET: 1185 Avenue of the Americas  
 ; CITY: New York  
 ; STATE: New York  
 ; COUNTRY: USA  
 ; ZIP: 10036  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patentin Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/821,818  
 ; FILING DATE:  
 ; CLASSIFICATION: 514  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Chan, Albert Wai-Kit  
 ; REGISTRATION NUMBER: 36,479  
 ; REFERENCE/DOCKET NUMBER: 51523  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (212) 278-0400  
 ; TELEFAX: (212) 391-0525  
 ; INFORMATION FOR SEQ ID NO: 3:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 657 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; US-08-821-818-3





```

Qy 597 ESQEDNCPGCGLSEALAGERYTHIKKVKVTFLEVTYYIS-----GDEDRKGWEEFAR 651
Db 448 -----PARNQGPFI-----LKGKVFHSEKVTVHFLAVWAGPAQAARRGPEQFAR 494
Qy 652 DGCERFQKRIQTEVAIGCYCLAFEHREKMFRLR 684
Db 495 DRSRFRARRIAQAEQGLGYLTPAPRARAWTLR 527

RESULT 6
US-08-893-852A-4
; Sequence 4, Application US/08893852A
; Patent No. 6080558
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Lal, Preeti
; APPLICANT: Shah, Purvi
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: NEW HUMAN GROWTH REGULATOR PROTEIN
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/893,852A
; FILING DATE: Filed Herewith
; PRIOR APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0341 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 590 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 452490
; US-08-893-852A-4

Query Match 4.6%; Score 171.5; DB 3; Length 590;
Best Local Similarity 18.1%; Pred. No. 8.1e-07;
Matches 100; Conservative 68; Mismatches 198; Indels 187; Gaps 18;

Qy 250 PESCYLSDGCHGPQLRAEMSATARRCPPL-----STEGLEPIHRR 292
Db 103 PETGLSDDDKQGDGPREQGRANTAGLIFLLSPGLQSAKSLGVVAGEGVTELAYPT 162
Qy 293 MRWLVLQPNQGQDLPTLDQNG-----YHSLREH 323
Db 163 SHW-----EGCPSEEDGETVKKAFRASADSPGHKSSTSVYCPGEAHOATEKQ 213
Qy 324 NLRMDP-----QCTDNPAQVSPADRP-----TEKKPELV 358
Db 214 TENKADPPSPSGSHSRAMEYCSKQEGA-----DPEPHRAGKYQLCQNAEAESEAK 267
Qy 359 IQEVSQSPQSSSL--FCLELPVEKECEDHTNATDLSDRGESLPVSTRPVCNKLIDYILG 416

```

```

Db 268 VSSLSVSSGNAFLKAWVYRPGEDTDDDDSGSAEEGKALSSPTSP-----HDPLKA 322
Qy 417 GAPSDLEASDSSESDWGEPEDDGDFDSGLSSESDVEODSEGLHLWNSFHSVDPYKPN 476
Db 323 WYVRPGEDTDDDDSGSAEE-----EGKALSSPTSPEDHFLKAW-----VVRPGE 369
Qy 477 FTATIQTAAARIAPRDPDSGTSWGSVCGVSGQEGPLPET-----516
Db 370 DT-----EDDQSDWGSAAKDGCLAQTATPHTSAFLKTWVCCPGEDTDDDC 416
Qy 517 -----PDHSGEEDDWEPS-----ADEAENKLWNSFCHSEDPYNLLNFKAPFPSPGK 564
Db 417 EVVVPEDSEAADPKSPSHEAQCLPGEQTEGL-----VEAHSLSFQVAFYLPGEKPA 470
Qy 565 NWKGRQDSKASSEVTVAFSGHHTLLSCKAQLLESQEDNCPGCGLSEALAGERYTHIKRK 624
Db 471 PW-----TAPKLPRLQRLTLRTPTQ---DQDPETP-----LRARK 505
Qy 625 VTFLEVTYYIS-----GDEDRKGWEEFARDCGRFQKRIQTEVAIGCYCLAFEHREK 679
Db 506 VHFSENVTVHFLAVWAGPAQAARRGPEQGLARDRSRFRARRIAQAEKGLGYLTPAFRA 565
Qy 680 FNRRLIESKDLLL 692
Db 566 WARLGNPSLPLAL 578

RESULT 7
US-08-821-818-2
; Sequence 2, Application US/08821818
; Patent No. 6146877
; GENERAL INFORMATION:
; APPLICANT: Fisher, Paul B.
; TITLE OF INVENTION: IDENTIFICATION OF THE PROGRESSION
; TITLE OF INVENTION: ELEVATED GENE-3 AND USES THEREOF
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/821,818
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Chan, Albert Wai-Kit
; REGISTRATION NUMBER: 36,479
; REFERENCE/DOCKET NUMBER: 51523
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 590 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-821-818-2

Query Match 4.6%; Score 171.5; DB 3; Length 590;
Best Local Similarity 18.1%; Pred. No. 8.1e-07;
Matches 100; Conservative 68; Mismatches 198; Indels 187; Gaps 18;

```



[illegible]

RESULT 10  
US-09-949-016-10933  
; Sequence 10933, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 10933  
; LENGTH: 3913  
; TYPE: PRT  
; ORGANISM: Human  
US-09-949-016-10933

Query Match 3.4%; Score 127.5; DB 4; Length 3913;  
Best Local Similarity 19.1%; Pred. No. 0.2;  
Matches 159; Conservative 113; Mismatches 346; Indels 211; Gaps 33;

Qy	5	THRARKPGRLGSRWFLPPLRSHACSEFFPPSSRQNFQNGSALPERRTRYWTKLLSQL	64
Db	1155	TFSSRTSPVTTAGS-----LLRSSITMTPPASPKSNINMYSSSLP-----FKSIITSA	1203
Qy	65	LALLPSLFO-----KLLWSQLS-----GGLPTRWLD	92
Db	1204	APLISSPLKVVSPVKSAVDVISAKITMASSLSPPVKQMPGHAEVALVNGSISPLKYP	1263
Qy	93	FAASYSALRASRGRESADPTVQKSLSYTAAGLFAKTRVYVSTLALARGGTPVAVLVRLE	152
Db	1264	SSTLINGCKATATLQEKISSATNSVSWSAATDVEKVFSTTT---AMPFSPSL-----	1314
Qy	153	VKLKAQERALDSPAFTLEQQLVGWELLPSLSQAGLVSHRELDSSSGPLSVQSLGNFK	212
Db	1315	-----RSVSAAPSAFQSLRTPSALYTSLGSSISA-----TSSVTSIIITVPVYS	1362
Qy	213	VWYLLNPSYLDLPQLGLRCQSSAGGQGVGFRTLTPESCYLSDECHQPQLRAEMSA-	271
Db	1363	VWVLPPEPA-LKKLPDSNSFTKAA--ALLSPIKTLTET-----HQPHPFSRTSSP	1411
Qy	272	-----TAWRRCPPLS-----TEGLPEIHRRMRWLVLQPNQGOQLP-----	308
Db	1412	VKSSFLAPSALLKSTPSSLSSSQELKQVAMKEKDLMRMTAILQTDVPEKFPQBELPK	1471
Qy	309	--TLDQDNGHSLSE--EHNLLR-----MDPOHCTDTPAQAVSPAADR--PETEKKPELVQ	360
Db	1472	EGRIDDEEPFKIVEKVKEDLVKVSLEILKDVCKDNKGPSKSPKSDKGHSPEDDWIEPSE	1531
Qy	361	EV-----SOSPQ-----GSSLFCFL-----P	376
Db	1532	EIREARQAAAASQSPSLPERVOVKAKAASEKYNLTKVIDYLTNDIGSSSLTNLYKPED	1591
Qy	377	VEKECEBDHTN-----ATDLSDRGESLP-----VSTRPVCSKNKLIYIILGGAPSDLEAS	425
Db	1592	AKKDGEERQKVLKPAITALQEHKLKMPASPMSRTSTSEKLC--KMADSPF-GTDTILESP	1648
Qy	426	SDSESDWGEEP--EDDGFSDGSLSESVDQDSEGLHLWNSFHSVDYKQNFATTIQT	484
Db	1649	DDFSQHDQDKSLSDSGFTTRSRTKTPS-AQCSAESTQPKPLFHEV-FIPPVITRTEV	1706
Qy	485	ARIAPRDPDSGTSWSGVCGVGCQGLPETPDHSSGEEDDWEPSADEAENKLMNSFC	544
Db	1707	HVIRSYDPSAGDVPQTQ-----PEEPVSPKSPPTMELEPKPTTSSIKEKVKAQWKA	1759
Qy	545	HS-EDPN-----LNFKAPFPQSGKNWGR-QDSKASSSVTVAFS	583
Db	1760	SSEDDHNRLVSLKGMVKEETHITTTTRMVYHSP--PGCGASERIEETMSVHDMKAFQ	1817
Qy	584	GHHTLLSCKAQLLESQEDNCGCGGLGALAGERYTHIKRKYVTLFBEVTEYISGDEDRK	643
Db	1818	SGRDPSEKLAGLFEHKSANVSPVHKSAAETSQAHAEDKQNKPKLERIEVHI-----	1872
Qy	644	GPWEEFARDGCRFKRIQIETEVAIGYCLAFEHREKMFNRLRIESKDLL	691
Db	1873	GNQAEPTVIRRETKKHPEKEMVY-----YQKLSRGDINLKDFL	1912
RESULT 11			
US-08-976-255-14			
; Sequence 14, Application US/08976255			
; Patent No. 6136581			
; GENERAL INFORMATION:			
; APPLICANT: Jono, Keith E.			
; APPLICANT: Plowman, Gregory			
; TITLE OF INVENTION: KINASE GENES AND USES			
; NUMBER OF SEQUENCES: 53			
; CORRESPONDENCE ADDRESS:			
; ADDRESSEE: Lyon & Lyon			
; STREET: 633 West Fifth Street			
; STREET: Suite 4700			
; CITY: Los Angeles			
; STATE: California			

RESULT 11  
US-08-976-255-14  
Sequence 14, Application US/08976255  
Patent No. 6136581  
GENERAL INFORMATION:  
APPLICANT: Jono, Keith E.  
APPLICANT: Plowman, Gregory  
TITLE OF INVENTION: KINASE GENES  
NUMBER OF SEQUENCES: 53  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lyon & Lyon  
STREET: 633 West Fifth Street  
STREET: Suite 4700  
CITY: Los Angeles  
STATE: California



```
;
; COUNTRY: U.S.A.
; ZIP: 90071-2066
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: FastSeq for Windows 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/976.255
; FILING DATE: No. 6136581ember 21, 1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/031.675
; FILING DATE: No. 6136581ember 22, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard J.
; REGISTRATION NUMBER: 321,327
; REFERENCE/DOCKET NUMBER: 229/182
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1503 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Protein
; US-08-976-255-14

Query Match          3.4%; Score 127; DB 3; Length 1503;
Best Local Similarity 21.4%; Pred. No. 0.049;
Matches 117; Conservative 55; Mismatches 204; Indels 172; Gaps 24;

Qy 171 LEQLMGVLL-----PSSIQAGLVSHRELDSSSGPLSVQSGNFKVSYLNPYSYLD 224
Dy ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 940 LEKNLEAVETLNQLNSKDAKEAGLVSAISDSSTQSLSLDS-----LSAPFPA 989

Qy 225 YLPOLGLRCQSSAGGQVGRFTITPESCYLESDGCHPQPLRAEMSATAMERCP----- 279
Dy ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 990 SEPSLE-----TPDS--LESVDVHEALDLSGSHTPQKLVPDPKPAD 1029

Qy 280 --LSTEGI--PEIHRRMRWLVLQPNQGD--PTLDQDNGYHSLE-----EEH 323
Dy ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1030 SGYETENLESP-----WTLHPAECTADSEPAATTCDDGCHSGLPNPNVIVISDAGDGH 1082

Qy 324 NLLRMDPQ-----HCTDNPQ-----AVSPAADPEPTEKKPELVIQ 360
Dy ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1083 RGTEVTPETFTAGSGYRDSAYFSDNDSEPEKSEEVPGTSPSALVLVQRPPLPEVLP 1142

Qy 361 EVSQSPQSSILFC--ELPVEKECEDHTNATDLSRGESLPVSTRPVCSNKLIDYILGAP 419
Dy ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1143 EQSPAQDSCUEARKSQPDESCLSAHNSDLELRATPEPAQT-----GVP 1188

Qy 420 SDLEASDSESEDW---GESPEPDGFD-----SDGSLSESDVEQDSEGLHLWNS 465
Dy ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1189 QQVHPTDEASSPMSVLNAELSSGDDPETQDDRPCTLASTGTNTNELLAYTNSALDKSL 1248

Qy 466 FHSVDYPKPNQFTATTIQTAAARIAPRDSGTSGVSGCGVSGCEGLPETPDHSSGEED 525
Dy ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1249 SHSEGP-----KLKEPDIEG-KYLGKLGV-----SGMLDLSDEGMDADEE 1287

Qy 526 DWEPSEDAEENLKLWNSFCHSEDPNLLNFKAPQPSGKNKWKGRQDSKASSEVTVAESGH 585
Dy ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1288 D-ENSDSDDEDLAFNLHSLSESEDETEHPVPVILNSD-----GR 1328

Qy 586 HTLSCKAQLLESQDNCPGCGGLAGEARYTHIKRKKVTFLEEVTEYIISGDEDRKGP 645
Dy ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1329 HL-----RSLLKPTAANAPD--PLPEDWKCE-----KKAVTFDDVTVILF----DQETP 1372

Qy 646 WEWFARDG 653
```

```
;
; COUNTRY: U.S.A.
; ZIP: 90071-2066
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: FastSeq for Windows 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/976.255
; FILING DATE: No. 6136581ember 21, 1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/031.675
; FILING DATE: No. 6136581ember 22, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard J.
; REGISTRATION NUMBER: 321,327
; REFERENCE/DOCKET NUMBER: 229/182
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1503 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Protein
; US-08-976-255-14

Query Match          3.4%; Score 125; DB 4; Length 2781;
Best Local Similarity 20.7%; Pred. No. 0.2;
Matches 99; Conservative 68; Mismatches 141; Indels 170; Gaps 27;

Qy 238 GGGQFVGFRITLPESCYLE-----DGCHPOPLRAEMSATAMRRCPPLSTEGLEPIH 289
Dy ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 947 GGIGKIG-KTSTNSSKNLSESPVITRAKEGCQSDSMRQEQS-----986

Qy 290 HRRMRWLVLQPNQGDPTLDQDNGYHSLEEHNLLEW--DPOHCTDNPQAQVSPAADRP 348
Dy ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 987 -----PNANDQP---EDLIQGCQSQSSSVLRMSDPSTHT-----NKL 1021

Qy 349 EPTKKPELVIOEVS--QSPQSSILFCFELPVEKECED--HTNATDLSRGESLPVSTRP-- 404
Dy ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1022 YPKDR---VLDDVSIKSP-----TKCPKQNSIENDIEBKVSDLASRGQE--PTKSKTKG 1071

Qy 405 -----VCSNK--LI-----DYILGAPSDLEASDSESDGEWGEDP 439
Dy ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1072 NDFFIDDSKLASADDITGLICKKXKPLIOESDTIVSSSKSALHSSVPKSTNDRDATPLS 1131

Qy 440 DGFDSQSL-----SESDVEQDSEGLHLWNSFHSVDPYKPNFTATTIQTAAARIAPRPS- 493
Dy ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1132 RAMDFEGKGCDSSESSTLNSSDTVSIQDS--SEEDMIVQNSNESISEQFRTREQDVEV 1189

Qy 494 -----DSGTSWSGCGVSGCEGLPETPDHSSGEEDDWEPSEDAE-----AENLKL 539
Dy ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1190 LEPLKELVSGES--TGNC-----EDRLPVKGTENGKPKSQKKLEERPVNKCSQDQIKL 1242

Qy 540 WNSFCHSEDPNLLNFK-----APPQSGKNKWKGRQDSKASSEVTVAESGHHTLLSC 591
Dy ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1243 KN-----TTDKNNENRESEKKQRTSTFQINGKDNKPK-----IYLGK-----EC 1283

Qy 592 KAQLLESQDNCPGCGGLAGEAR---RYTHIKR-----KKVTFLEEVTEYIISGD 639
Dy ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1284 LKEISESR-----VSGNVEPKVNNINKIPENDIKSLTVKESAIRPFINGD 1330

RESULT 13
US-09-698-295-1
; Sequence 1, Application US/09698295
; Patent No. 6689584
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATORY FACTOR
; FILE REFERENCE: 06501-068001
; CURRENT APPLICATION NUMBER: US/09/698,295
; CURRENT FILING DATE: 2000-10-27
```



1094	KVDAQEAKTPTFTGKVVGGQTTPSFKAQOVTSIESSELVUWTCQAEITLAGVKSQEMVM	115
408	NKLIIDYILGAPSDLEASSDESDESDWCEEPEDDGFDSGLSESDVEQDSGLHLWNSFH	467
1154	EQAI-----PPDSVETPTDSETD--GSTPVAD--FDAPGTTQKDEIVE---IHEENEVH	1200
468	SVD-----PYKQNF-----ATIOFAARIAPRD	491
1201	LVPVRGTEABVPAQKERPPAPSPFVFOETKEOSKMEDTLEHTDKEVSVETVLSLSTE	1266
492	PSDSGTSWS-----GSCGVG-----SCQSGPLPET	516
1261	GTQEAQYADEKTKDVFFPSGLESGIDTIGITVSRKEVTEVALKCGETEEAECKDDALEL	1320
517	PDHSSG-----BEDDWPSADAENLKLWNSFCHSEPDYN--LLNFKAPP	559
1321	QSHAKSPSPVERMMVVQVREKTEAEPTHNEEKLESHETAVTVSEEVSKOLLTOVNVPI	1380
560	QPSGK---NWKGR-----QDSKASSEVTVAF-----	582
1381	IDGAKEVSSLEGGPPCLGQEEAVCTKIQVSSSEASFTLTAABSEKVLGETANILETGE	1440
583	-----SGHHTLLSKCAQLLESQED-----NCPGC-----GLGEAL	612
1441	TLBPAGAHVLVEBKSS--EKNEDEFAAHPGEDAVPTGPDCAKSTPVIVSATTKKGLSSDL	1498
613	AGRYRTHIKRKVTFFLEEVTEYYISGDEDRKGPWEFPARDRCRQKRIQETEAIVGCLA	672
1499	EGETTSLKWS-----DEVDE-----QVACQEVKVSU---A	1522
673	FEHREKMFNRLRIESKDLLLSYNV	696
1528	IEDLEPENGILETETSSKIVQNI	1551

Search completed: September 16, 2005, 10:47:19  
Job time : 20.798 secs





```

Qy 350 PTEKKPELVIOEV-----SQSPQSSSLFCELPVKECEEDHTNATDLSRGSLSLVSTRP 404
Db 360 STEEKIELLTTEVPLALEEESPSCPSSEIPMEKBPCEGRISVVDYSYLEGDLPIGARP 419
Qy 405 VCSNKLIDYILGGAPSLDLEASDSSESDEWGEPEDDGFDSDGSLSESDVEQDSGLHWN 464
Db 420 ACSNKLIDYILGGASSDLETSDDGEDEWDEADDGFDSDSSLSDSLEQDPGGLHWN 479
Qy 465 SFHSVDPYKPNFTATTIQTAAARIAPRDPDSGTSWSSGCV-GSCQEGPLPETPDHSSGE 523
Db 480 SFCSDVPYKPNFTATTIQTAAARIAPRDPDSGTSWSSGCV-GSCQEGPLPETPDHSSGE 539
Qy 524 EDDPEPSADEAKNLWNSFCHSDPYNLNFKAPFPQSGKWKGRDSDKASSEVTVAPS 583
Db 540 EDDWESSADEABSLLWNSFCNSDDPYNPLNFKAPFPQSGKWKGRDSDKTPSPSSIVAIS 599
Qy 584 GHHTLLSCKAQLLESQEDNCGCGLGEALAGERVTHIKRKKVTFLEEVTEYIISGDEDRK 643
Db 600 ECHTLLSCKVQLLSQSECEPDSVQDVLSGRTHVKKVTFLEEVTEYIISGDEDRK 659
Qy 644 GPWEEFARDGCRFQKRIQETEAIGYCLAFEHREKMFNRLR 684
Db 660 GPWEEFARDGCRFQKRIQETEAIGYCLTFEHRERMFNRLQ 700

RESULT 3
US-09-925-299-941
; Sequence 941, Application US/09925299
; Patent No. US20020055627A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 941
; LENGTH: 707
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (265)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (271)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (307)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-299-941

Query Match 58.2%; Score 2157.5; DB 9; Length 707;
Best Local Similarity 64.2%; Pred. No. 2.7e-156; Indels 17; Gaps 6;
Matches 442; Conservative 54; Mismatches 176; Indels 17; Gaps 6;

Qy 12 PGPLGSLFRLPFLRRSHACSEPPPPSSRRQPN-----SALPERRTRYWTKLLSLLA 66
Db 7 PVSRCSGAQFQSVRRSQAGSKFPTPLGPENSGNPTLLSSAQPTRVSVYWKLLSLLA 66
Qy 67 LLPSLFQKLLWSQSLGGLIETRMDFPAASYSALRASGREESAPTVQKSLSVTAAGLF 126
Db 67 PLPLGLQKVLWSQSLFGGMFTRMDFAGVYSALRALKAGREKPAAPTAQKSLSLQLD-S 125
Qy 127 AKTRVSVTLALRGGTPVAVLVLEVKLKAQERALDSAAPTFLLEQOLWGVELLPSSLQ 186
Db 126 SDPSVTSPLDWLEBEGHWHQVSPFDLKLKELKAKGALDPAQAFLLEQOLWGVELLPSSLQ 185

```

```

Qy 187 AGLVSHRELDSSSGPLSVQSLGNFKVSVLLNPSYLDYLPOLGLRCSOAGGQFVGPR 246
Db 186 SRLYSNRELSSPSGLNIQRIIDNFVSVYLLNPSYLDLCPRLLEVSYQNSDGNSEVVGFG 245
Qy 247 TLTPESSCYLSEDCGHPOLRAEMSAVARRCPPLSTEGLEPEIHHRRMRWLVL-OPNQGG 305
Db 246 TLTPESSCLREDCHQPLXAEILIPXSWQCCPLSTEGLEPEIHHLRMKRLEFLQOASKGQ 305
Qy 306 DLFTLDQNGYHLSLEBEHNLLRMDPQHCNTDNPAAQAVSPAADRP-----EPTKKPELVIOE 361
Db 306 DXPTPDQDNGYHLSLEBEHSLLRMDPKHCRDNPQFVPAAGDIIPGNTQESTEEKIELTTE 365
Qy 362 V-----SQSPQSSSLFCELPVKECEEDHTNATDLSRGSLSLVSTRPVCNKLIDYILG 416
Db 366 VPLALEEESPSCPSSEIPMEKBPCEGRISVVDYSYLEGDLPISARPAACSNKLIDYILG 425
Qy 417 GAPSDLEASDSSESDEWGEPEDDGFDSDGSLSESDVEQDSGLHWNSEFSDVPYKPN 476
Db 426 GASDLETSSDPGEDWDEAEDGFDSDSLSDLEQDPGGLHWNSEFSDVPYKPN 485
Qy 477 FTATIQTAARIAPRDPDSGTSWSSGCV-GSCQEGPLPETPDHSSGEEDDWPSADEAE 535
Db 486 FTATIQTAARIAPRDPDSGTSWSSGCV-GSCQEGPLPETPDHSSGEEDDWPSADEAE 545
Qy 536 NLKLWNSFCHSDPYNLNFKAPFPQSGKWKGRDSDKASSEVTVAPSGHHTLLSCKAQL 595
Db 546 SLKLWNSFCNSDDPYNPLNFKAPFPQSGKWKGRDSDKASSEVTVAPSGHHTLLSCKAQL 605
Qy 596 LESQEDNCGCGLGEALAGERVTHIKRKKVTFLEEVTEYIISGDEDRKGPWEEFARDGCR 655
Db 606 LGSQSECPDSVQDVLSGRTHVKKVTFLEEVTEYIISGDEDRKGPWEEFARDGCR 665
Qy 656 FQKRIQETEAIGYCLAFEHREKMFNRLR 684
Db 666 FQKRIQETEAIGYCLTFEHRERMFNRLQ 694

RESULT 4
US-09-925-299-941
; Sequence 941, Application US/09925299
; Publication No. US20030040617A9
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 941
; LENGTH: 707
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (265)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (271)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (307)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-299-941

Query Match 58.2%; Score 2157.5; DB 10; Length 707;
Best Local Similarity 64.2%; Pred. No. 2.7e-156; Indels 17; Gaps 6;
Matches 442; Conservative 54; Mismatches 176; Indels 17; Gaps 6;

```





```

Db 479 ---LQPORCLPGEKTKGRGEEPLFQVAFYLPCEKPEPWAAPKPLRLQRLRLFKAPT 535
Qy 599 --QEDNCPGCGLGEALAGERYTHIKKKVTFLEEVTEYIIS-----GDEDKGPWESEFAR 651
Db 536 RDQDPEIP-----LQPORCLPGEKTKGRGEEPLFQVAFYLPCEKPEPWAAPKPLRLQRLRLFKAPT 535
Qy 652 DGRFQKRIQETEAIGYCLAFEHREKMFNRLR 684
Db 581 DRSRFAIRIAQAEKLGPLYLTPDSRARAWARLR 613

RESULT 7
US-10-277-603-3
; Sequence 3, Application US/10277603
; Publication No. US20030219376A1
; GENERAL INFORMATION:
; APPLICANT: Fisher, Paul B.
; TITLE OF INVENTION: PROGRESSION ELEVATED GENE-3 AND USES
; FILE REFERENCE: A34608-B-A
; CURRENT APPLICATION NUMBER: US/10/277,603
; CURRENT FILING DATE: 2002-10-22
; PRIOR APPLICATION NUMBER: 09/052,753
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: PCT/US98/05793
; PRIOR FILING DATE: 1998-03-20
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 657
; TYPE: PRT
; ORGANISM: Rat
US-10-277-603-3

```

```

Query Match 5.5%; Score 204; DB 15; Length 657;
Best Local Similarity 21.5%; Pred. No. 1.5e-06;
Matches 149; Conservative 86; Mismatches 290; Indels 168; Gaps 27;

Qy 60 LLSQALLPSLFQKLLWSLQSGGLIPTRWLDPAASYSALRASRGRESDAPTQVKSLS 119
Db 21 LLSPLMGLLR-----AWSRLR-GLGPLEPWLVEAVKGAAL----- 55
Qy 120 YTAAGLFAKTRVSTLALARG-----GTPVAVLVLRLEVKLKAQERALDSAAP-----T 168
Db 56 IERAAULTPTVPVGNL-LPHGETESGSP-----EQQAQRLCLVEAESPP----- 102
Qy 172 EQQLWGVLLPS-SLQAGLVSHRELD-SSSSGPLSVQSLG-----NFKVSY-LLN 219
Db 103 --ETWGLSNVDEYNAKPGQDDLREKEMERTAGKATLQAGLQAGADKRLGEVAREEGVAE 160
Qy 220 PSYLDYLPQLGLRCQSSAGGQFVGFRTLTPESCYLSGDPCHPQPLRAEWSATWRRCP 279
Db 161 PAY-----PTSOLEGPAEN-----BEDGETVITYQASAASTAPGVKPS 199
Qy 280 LSTEGLEPIHRRMRWLVLQPNQGLDPLTDQNGYHSLSEELHLLRMDPQHCTDNPAQ 339
Db 200 TPVPFLGEAEHQATE-----EKGTEKADPNPSGSGSHSRAWEY-YREKPKQGEAKVE 254
Qy 340 AVSPAARDP-----EPTKPKPELVITQEVSPQSGSLFCELPVEKCEBDHTNATDLSRG 395
Db 255 AHRAGQGHPCRNAAEAGGPEPTTP--VCTGNAFKAWVYRPGEDTEEDNSDSAEEDT 312
Qy 396 ESLPVSTRPVCNSKLIIDYILGAPSDLEASDSEED-----W-----GE 435
Db 313 AQTGATHTSAFLKAWVYRPGEDTEEDSDSAEEDTAQTGATPHYSAFLKAWVYRPG 372
Qy 436 EPEDGDGDSGLSESDVEQDSEGLHLSNFSHSDVPYKPNFTATIQTAARIAPRPSDS 495
Db 373 DTEENSDDLDS--AEBDTAQTGATHT-SAFKAWVYRPGEDTEEDNSDLDSAEEDTAQT 429
Qy 496 GTSMGSGCGVSGCOEGPLPPTPHSSGEEDDWEPSADEANLKLWNSFCHSDPYNLNLF 555

```

```

Db 430 GATPHTSPFLKAWVYRPGEDTDEEDSEENAVPGDSETADSSQSPC----- 478
Qy 556 KAPFQPS-----GNWKGRQDSKASSEVTVAFSGH-----TLLSCKAQLLES-- 598
Db 479 ---LQPORCLPGEKTKGRGEEPLFQVAFYLPCEKPEPWAAPKPLRLQRLRLFKAPT 535
Qy 599 --QEDNCPGCGLGEALAGERYTHIKKKVTFLEEVTEYIIS-----GDEDKGPWESEFAR 651
Db 536 RDQDPEIP-----LQPORCLPGEKTKGRGEEPLFQVAFYLPCEKPEPWAAPKPLRLQRLRLFKAPT 535
Qy 652 DGRFQKRIQETEAIGYCLAFEHREKMFNRLR 684
Db 581 DRSRFAIRIAQAEKLGPLYLTPDSRARAWARLR 613

RESULT 8
US-10-755-889-90
; Sequence 90, Application US/10755889
; Publication No. US20040171823A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES ASSOCIATED WITH THE NF-kB
; FILE REFERENCE: D0284 NP
; CURRENT APPLICATION NUMBER: US/10/755,889
; CURRENT FILING DATE: 2004-01-13
; PRIOR APPLICATION NUMBER: U.S. 60/440,068
; PRIOR FILING DATE: 2003-01-14
; PRIOR APPLICATION NUMBER: U.S. 60/469,757
; PRIOR FILING DATE: 2003-05-12
; NUMBER OF SEQ ID NOS: 823
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 90
; LENGTH: 674
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-755-889-90

```

```

Query Match 5.0%; Score 186.5; DB 16; Length 674;
Best Local Similarity 21.5%; Pred. No. 3.4e-05;
Matches 156; Conservative 73; Mismatches 270; Indels 227; Gaps 32;

Qy 60 LLSQALLPSLFQKLLWSLQSGGLIPTRWLDPAASYSALRASRGRESDAPTQVKSLS 118
Db 21 LLSPLMGLLR-----AWSRLR-GLGPLEPWLVEAVKGAAL----- 55
Qy 119 SYTAAGLFAKTRVSTLALARGGTPVAVLVLRLEVKLKAQERALDSAAP-----T 168
Db 56 --VEAGLEGEAR-----TPLA--IPHTPWGRRPGEEAEDSGGPGEDRETGLKLT 100
Qy 169 FLLEQQLWGVLLPSLQAGLVSHRELDSSSSGPLSVQSLGNFKVSYLLNPSYL----- 223
Db 101 SSSLPEAWG--LLDD--DDGWYGEREATSVPRGSGQFADGQ---RAPLSPLLIRTLQ 152
Qy 224 --DYLQGLRLCQSSAGGQFVGFRTLT-----PSCYLSGDPCHPQPLRAEWSATA--- 273
Db 153 GSDKNP--GSEKAEGBEAGVAAEENKFPSPHRECCPAVEEEDDEAVKKEAHTSTSA 210
Qy 274 -----WRRCP-----PLSTEGLEPIHRRMRWLVLQPNQGLDPLTDQNGYHSL 320
Db 211 LSPGSKPSTWVSCFGBEENQATDKKTERSKGARKTSVSPRSGSDPRSWEYSGASEE 270
Qy 321 EEHLLRMDPQHCTDNPAQVSPAARDPTEKPKPELVITQEVSPQSGSLFCELPVEKE 380
Db 271 KE-----EKAHETKGEAA-----PGQSSAP-----AQRPLKSMWCQPSDEE 311
Qy 381 CREDHTNATDLSRGSLPVSTRPVCNSKLIIDYILGAPSDLEASDSEEDGCEPEDD 440
Db 312 SEVKPLGAAEKQGEAECCPCIPPPSAFLKAWVYWPWPG---EDTEEEDEE---DESDS 364
Qy 441 GPDSDGSLSESDVEQDSEGLHL---WNSFHSVDYKPNFTATIQ-----TAARIAPRPS 493
Db 365 GSDEEGBEAEASSTPATGVFLKSW-----VIQPGEDTEEDDESDTGSADEREAE 417

```





```

; FILE OF INVENTION: 38-21(5313)B
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 66899
; LENGTH: 830
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: UC-2MFLMOL7068A01_FLI.pep
; US-10-425-114-66899

Query Match          4.1%; Score 152.5; DB 15; Length 830;
Best Local Similarity 21.5%; Pred. No. 0.018;
Matches 90; Conservative 59; Mismatches 133; Indels 137; Gaps 22;

Qy 307 LPTLDQ-----DNGYHSL-----EEHNNLLRMD---PQCHTDNPAQVSPAADRPETKEKP 355
Db 492 LPAMSKSLMWDNRSLUKVQWDFSSLLRPDRTPWEV--EPDLAANPQSGQPPLRAKRPP 549
Qy 356 ELVIOE--VSQSPQGGSLFCELVIEKECEBEDHTNATDLSDRGSLPVSTRPVCSNKLIDY 413
Db 550 RPPASPCMWSELPSGGFLM--KSPIESCTLSFGEPPQARELFPSIPTSTLSSEN-----603

```



**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 16, 2005, 10:40:09 ; Search time 178.591 Seconds  
(without alignments)  
4565.254 Million cell updates/sec

Title: US-10-650-482-4

Perfect score: 3710

Sequence: 1 METGTHARKRPGRLGSWF.....MFNRLRIESKLLLYSNVKK 698

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 6959266 seqs, 1168006243 residues

Total number of hits satisfying chosen parameters: 6959266

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

Pending Patents AA Main:\*

1: /cgn2\_6/ptodata/1/paa/US06 COMB.pcp.\*  
2: /cgn2\_6/ptodata/1/paa/US06 COMB.pcp.\*  
3: /cgn2\_6/ptodata/1/paa/US07 COMB.pcp.\*  
4: /cgn2\_6/ptodata/1/paa/US08 COMB.pcp.\*  
5: /cgn2\_6/ptodata/1/paa/US08 COMB.pcp.\*  
6: /cgn2\_6/ptodata/1/paa/US08 COMB.pcp.\*  
7: /cgn2\_6/ptodata/1/paa/US08 COMB.pcp.\*  
8: /cgn2\_6/ptodata/1/paa/US08 COMB.pcp.\*  
9: /cgn2\_6/ptodata/1/paa/US08 COMB.pcp.\*  
10: /cgn2\_6/ptodata/1/paa/US08 COMB.pcp.\*  
11: /cgn2\_6/ptodata/1/paa/US08 COMB.pcp.\*  
12: /cgn2\_6/ptodata/1/paa/US08 COMB.pcp.\*  
13: /cgn2\_6/ptodata/1/paa/US08 COMB.pcp.\*  
14: /cgn2\_6/ptodata/1/paa/US09 COMB.pcp.\*  
15: /cgn2\_6/ptodata/1/paa/US09 COMB.pcp.\*  
16: /cgn2\_6/ptodata/1/paa/US09 COMB.pcp.\*  
17: /cgn2\_6/ptodata/1/paa/US09 COMB.pcp.\*  
18: /cgn2\_6/ptodata/1/paa/US09 COMB.pcp.\*  
19: /cgn2\_6/ptodata/1/paa/US09 COMB.pcp.\*  
20: /cgn2\_6/ptodata/1/paa/US09 COMB.pcp.\*  
21: /cgn2\_6/ptodata/1/paa/US09 COMB.pcp.\*  
22: /cgn2\_6/ptodata/1/paa/US09 COMB.pcp.\*  
23: /cgn2\_6/ptodata/1/paa/US09 COMB.pcp.\*  
24: /cgn2\_6/ptodata/1/paa/US09 COMB.pcp.\*  
25: /cgn2\_6/ptodata/1/paa/US09 COMB.pcp.\*  
26: /cgn2\_6/ptodata/1/paa/US10 COMB.pcp.\*  
27: /cgn2\_6/ptodata/1/paa/US10 COMB.pcp.\*  
28: /cgn2\_6/ptodata/1/paa/US10 COMB.pcp.\*  
29: /cgn2\_6/ptodata/1/paa/US10 COMB.pcp.\*  
30: /cgn2\_6/ptodata/1/paa/US10 COMB.pcp.\*  
31: /cgn2\_6/ptodata/1/paa/US10 COMB.pcp.\*  
32: /cgn2\_6/ptodata/1/paa/US10 COMB.pcp.\*  
33: /cgn2\_6/ptodata/1/paa/US10 COMB.pcp.\*  
34: /cgn2\_6/ptodata/1/paa/US10 COMB.pcp.\*  
35: /cgn2\_6/ptodata/1/paa/US10 COMB.pcp.\*  
36: /cgn2\_6/ptodata/1/paa/US10 COMB.pcp.\*  
37: /cgn2\_6/ptodata/1/paa/US60 COMB.pcp.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	3710	100.0	698	32	US-10-650-482-4	Sequence 4, Appl
2	2223	59.9	713	32	US-10-650-482-2	Sequence 2, Appl
3	2218	59.8	713	37	US-10-170-205E-23916	Sequence 23916, A
4	2218	59.8	713	37	US-10-650-482-680-17655	Sequence 17655, A
5	2213	59.6	713	37	US-09-629-469A-18965	Sequence 18965, A
6	2213	59.6	713	35	US-10-917-503-18965	Sequence 18965, A
7	2208	59.5	720	22	US-09-757-028-2348	Sequence 2348, Ap
8	2208	59.5	720	28	US-10-222-911-2348	Sequence 2348, Ap
9	2157.5	58.2	707	1	PCT-US00-05883-941	Sequence 941, App
10	2157.5	58.2	707	24	US-09-925-299-941	Sequence 941, App
11	547	14.7	153	1	PCT-US00-35017A-963	Sequence 963, App
12	547	14.7	153	28	US-10-296-115-963	Sequence 15743, A
13	357	9.6	101	1	PCT-US01-14827-15743	Sequence 15743, A
14	307	8.3	67	37	US-60-196-718-7356	Sequence 7356, Ap
15	307	8.3	73	37	US-60-177-571-4491	Sequence 4491, Ap
16	304	8.2	73	37	US-60-162-247-4518	Sequence 4518, Ap
17	304	8.2	73	37	US-60-169-840-8186	Sequence 8186, Ap
18	204	5.5	657	12	US-08-893-852-3	Sequence 3, Appl
19	204	5.5	657	28	US-10-274-666-3	Sequence 3, Appl
20	204	5.5	657	28	US-10-277-603-3	Sequence 1, Appl
21	191.5	5.2	674	12	US-08-893-852-1	Sequence 905, App
22	191.5	5.2	697	34	US-10-821-234-905	Sequence 18860, A
23	189.5	5.1	674	27	US-10-170-205E-18860	Sequence 549, App
24	189.5	5.1	674	28	US-10-219-051B-549	Sequence 553, App
25	189.5	5.1	674	28	US-10-219-051B-553	Sequence 10552, A
26	189.5	5.1	674	28	US-10-219-051B-10552	Sequence 10552, A
27	189.5	5.1	674	37	US-60-443-566-4148	Sequence 4148, Ap
28	189.5	5.1	674	37	US-60-452-680-23269	Sequence 23269, A
29	189.5	5.1	674	37	US-60-453-050-14551	Sequence 14551, A
30	189.5	5.1	674	37	US-60-453-135-14551	Sequence 14551, A
31	189.5	5.1	674	37	US-60-455-444-7838	Sequence 7838, Ap
32	189.5	5.1	674	37	US-60-465-241-7838	Sequence 7838, Ap
33	189.5	5.1	674	37	US-60-466-412-14551	Sequence 14551, A
34	186.5	5.0	674	20	US-09-629-469A-11494	Sequence 11494, A
35	186.5	5.0	674	33	US-10-755-889-90	Sequence 90, Appl
36	186.5	5.0	674	35	US-10-917-503-11494	Sequence 11494, A
37	186.5	5.0	674	37	US-60-440-068-90	Sequence 90, Appl
38	186.5	5.0	674	37	US-60-469-757-90	Sequence 90, Appl
39	182.5	4.9	578	16	US-09-236-772-2	Sequence 2, Appl
40	182.5	4.9	578	28	US-10-274-666-7	Sequence 7, Appl
41	182.5	4.9	578	28	US-10-277-603-7	Sequence 4, Appl
42	171.5	4.6	590	12	US-08-893-852-4	Sequence 2, Appl
43	171.5	4.6	590	28	US-10-274-666-2	Sequence 2, Appl
44	171.5	4.6	590	28	US-10-277-603-2	Sequence 55843, A
45	156	4.2	627	21	US-09-724-676-55843	Sequence 55843, A
46	156	4.2	627	21	US-09-724-676A-55843	Sequence 31529, A
47	152.5	4.1	612	27	US-10-155-881-31529	Sequence 38378, A
48	152.5	4.1	612	30	US-10-425-114-38378	Sequence 38378, A
49	152.5	4.1	612	30	US-10-425-114A-38378	Sequence 17030, A
50	152.5	4.1	612	37	US-60-324-109-17030	Sequence 207028, A
51	152.5	4.1	661	30	US-10-425-115-207028	Sequence 5980, Ap
52	152.5	4.1	661	30	US-10-438-246-5980	Sequence 58227, A
53	152.5	4.1	830	28	US-10-219-999-58227	Sequence 66899, A
54	152.5	4.1	830	30	US-10-425-114-66899	Sequence 66899, A
55	152.5	4.1	830	30	US-10-425-114A-66899	Sequence 27894, A
56	152.5	4.1	830	37	US-60-324-109-27894	Sequence 7333, Ap
57	152.5	4.1	834	27	US-10-155-881-7333	Sequence 152, App
58	149.5	4.0	923	1	PCT-US02-10780-152	Sequence 1354, Ap
59	149.5	4.0	923	27	US-10-114-270-152	Sequence 1354, Ap
60	149	4.0	1365	1	PCT-US03-40978-1354	Sequence 704, App
61	149	4.0	1365	33	US-10-741-600-1354	Sequence 439, App
62	149	4.0	1365	33	US-10-796-307-704	Sequence 439, App
63	149	4.0	1365	35	US-10-995-561-867	Sequence 13533, A
64	149	4.0	1365	37	US-60-568-219-439	Sequence 13533, A
65	148.5	4.0	676	20	US-09-614-150-13533	Sequence 11065, A
66	148.5	4.0	676	20	US-09-614-150A-13533	Sequence 11065, A
67	148.5	4.0	676	22	US-09-791-537-11065	Sequence 13595, A
68	148.5	4.0	676	37	US-60-167-217-13595	Sequence 11069, A
69	148.5	4.0	676	37	US-60-173-464-11069	



```

70 148.5 4.0 676 37 US-60-191-637-13574 Sequence 13574, A
71 148.5 4.0 676 37 US-60-191-681-10700 Sequence 10700, A
72 147.5 4.0 725 20 US-09-629-469A-17386 Sequence 17386, A
73 147.5 4.0 725 35 US-10-917-503-17386 Sequence 17386, A
74 147.5 4.0 728 27 US-10-170-205B-22564 Sequence 22564, A
75 147.5 4.0 728 37 US-60-452-680-17245 Sequence 17245, A
76 147.5 4.0 728 37 US-60-453-050-10486 Sequence 10486, A
77 147.5 4.0 728 37 US-60-453-135-10486 Sequence 10486, A
78 147.5 4.0 728 37 US-60-456-412-10486 Sequence 10486, A
79 147.5 4.0 835 27 US-10-170-205E-23688 Sequence 23688, A
80 147.5 4.0 835 33 US-10-755-889-36 Sequence 36, Appl
81 147.5 4.0 835 37 US-60-440-068-36 Sequence 36, Appl
82 147.5 4.0 835 37 US-60-452-680-17244 Sequence 17244, A
83 147.5 4.0 835 37 US-60-453-050-10485 Sequence 10485, A
84 147.5 4.0 835 37 US-60-453-135-10485 Sequence 10485, A
85 147.5 4.0 835 37 US-60-466-412-10485 Sequence 10485, A
86 147.5 4.0 835 37 US-60-469-757-36 Sequence 36, Appl
87 144 3.9 867 1 PCT-US01-29288A-130 Sequence 130, Appl
88 144 3.9 867 1 PCT-US01-29288A-130 Sequence 130, Appl
89 144 3.9 867 18 US-09-487-558-98 Sequence 98, Appl
90 144 3.9 867 18 US-09-487-558-98 Sequence 98, Appl
91 144 3.9 867 22 US-09-751-537-92178 Sequence 92178, A
92 144 3.9 867 27 US-09-801-368-98 Sequence 98, Appl
93 144 3.9 867 27 US-10-149-310-130 Sequence 130, Appl
94 142.5 3.8 917 1 PCT-US04-24424-1542 Sequence 1542, Ap
95 142.5 3.8 917 37 US-60-490-890-1542 Sequence 1542, Ap
96 141 3.8 917 1 PCT-US02-26322-15 Sequence 15, Appl
97 141 3.8 917 30 US-10-487-092-15 Sequence 15, Appl
98 141 3.8 1198 33 US-10-777-288A-3349 Sequence 3349, Ap
99 139.5 3.8 883 1 PCT-US02-22539-6 Sequence 6, Appl
100 139.5 3.8 883 1 PCT-US03-17500-6 Sequence 6, Appl

```

# ALIGNMENTS

```

RESULT 1
US-10-650-482-4
; Sequence 4, Application US/10650482
; GENERAL INFORMATION:
; APPLICANT: Ron, David
; TITLE OF INVENTION: METHODS OF SCREENING TEST COMPOUNDS USING GADD34L, AN eIF2alpha-S
; TITLE OF INVENTION: PHOSPHATASE SUBUNIT
; FILE REFERENCE: 5986/11712-US1
; CURRENT APPLICATION NUMBER: US/10/650,482
; CURRENT FILING DATE: 2003-08-28
; PRIOR FILING DATE: 2002-09-06
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 698
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-650-482-4

```

```

Query Match 100.0%; Score 3710; DB 32; Length 698;
Best Local Similarity 100.0%; Pred. No. 1e-313;
Matches 698; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 METGTHARKKPGPRIGSWFLPFLRRSHACSSBPPPPSSRQNPNSALPERRTRYWTKL 60
Db 1 METGTHARKKPGPRIGSWFLPFLRRSHACSSBPPPPSSRQNPNSALPERRTRYWTKL 60

Qy 61 LSQALLPSLFOKLLWSQSGGLIPTRWLDFAASYSALRASGRRESADPTVOKLSY 120
Db 61 LSQALLPSLFOKLLWSQSGGLIPTRWLDFAASYSALRASGRRESADPTVOKLSY 120

Qy 121 TAAGLFAKTRVVTALARGGTPVAVLVLRLEVKLKAERALDSAAPTFLLEQQWGVEL 180
Db 121 TAAGLFAKTRVVTALARGGTPVAVLVLRLEVKLKAERALDSAAPTFLLEQQWGVEL 180

```

```

181 LPSLQAGLVSHRELDSSSGPLSVQSLGNFKVSVYLLNPSTYLDLPQLGLRCOSSAGG 240
181 LPSLQAGLVSHRELDSSSGPLSVQSLGNFKVSVYLLNPSTYLDLPQLGLRCOSSAGG 240
241 QFVGFRLLTPESCVLSYSDGCHPOPLRAEMGATARRCPPLSTEGLPRIHRRMRWLVLQ 300
241 QFVGFRLLTPESCVLSYSDGCHPOPLRAEMGATARRCPPLSTEGLPRIHRRMRWLVLQ 300
301 PNOQDLPPTLDQONGVHSLSEHNLLRMDPQHCTDNPQAQVSPAADPEPTEKKPELVQ 360
301 PNOQDLPPTLDQONGVHSLSEHNLLRMDPQHCTDNPQAQVSPAADPEPTEKKPELVQ 360
361 EVSPOGSSSLFCPLPEKECEEDHTNATDLSRGSPLVSTRPVCNSKLDIYILGAPS 420
361 EVSPOGSSSLFCPLPEKECEEDHTNATDLSRGSPLVSTRPVCNSKLDIYILGAPS 420
421 DLEASSDSESDGEEPEDDGFDGSLSDVEQDSEGLHLMNSFHSVDPYKQNTAT 480
421 DLEASSDSESDGEEPEDDGFDGSLSDVEQDSEGLHLMNSFHSVDPYKQNTAT 480
481 IOTAAARTAPRDPDSGTSWSGSCVGSQCGPLPETPDHSGSREDDWEPADAEENIKLW 540
481 IOTAAARTAPRDPDSGTSWSGSCVGSQCGPLPETPDHSGSREDDWEPADAEENIKLW 540
541 NSFCHSEDPYNLLNFKAFFQPSGKNWKGRODSKASSEVTVAFGHHTLLSCKAQLLESQ 600
541 NSFCHSEDPYNLLNFKAFFQPSGKNWKGRODSKASSEVTVAFGHHTLLSCKAQLLESQ 600
601 DNPFGCGLGALAGERYTHIKRKKVTFLEVTYIYISGDEDRKGPWEFPARDGCRFOKRI 660
601 DNPFGCGLGALAGERYTHIKRKKVTFLEVTYIYISGDEDRKGPWEFPARDGCRFOKRI 660
661 QETEVATGYCLAFEHREKMFENRLRIESKOLLVSNVKK 698
661 QETEVATGYCLAFEHREKMFENRLRIESKOLLVSNVKK 698

```

## RESULT 2

```

US-10-650-482-2
; Sequence 2, Application US/10650482
; GENERAL INFORMATION:
; APPLICANT: Jousse, Celine
; TITLE OF INVENTION: METHODS OF SCREENING TEST COMPOUNDS USING GADD34L, AN eIF2alpha-S
; TITLE OF INVENTION: PHOSPHATASE SUBUNIT
; FILE REFERENCE: 5986/11712-US1
; CURRENT APPLICATION NUMBER: US/10/650,482
; CURRENT FILING DATE: 2003-08-28
; PRIOR FILING DATE: 2002-09-06
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 713
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-650-482-2

```

```

Query Match 59.9%; Score 2223; DB 32; Length 713;
Best Local Similarity 65.0%; Pred. No. 7.8e-184;
Matches 456; Conservative 53; Mismatches 174; Indels 18; Gaps 7;

Qy 1 METGTHARKKPGPRIGSWFLPFLRRSHACSSBPPPPSSRQNPNGN-----SALPERRT 54
Db 1 MEFGTGSRRKRLPFRAGFRFWPPFFPFRSOGSKFPTPLGPENSGNPTLLSSAQPETRV 60

Qy 55 RYWTKLLSQALLPSLFOKLLWSQSGGLIPTRWLDFAASYSALRASGRRESADPTV 114
Db 61 SYWTKLLSQALLPSLFOKLLWSQSGGLIPTRWLDFAAGVTSALRALGRKPKAPTA 120

Qy 115 QKSLSYTAAGLFAKTRVVTALARGGTPVAVLVLRLEVKLKAERALDSAAPTFLLEQQ 174
Db 121 QKSLSYQLD--SSDPSVTSPLDMLWLEGIHWQYSPDPLKLELKAGSALDPAQAFLLEQQ 179

```

180 LMWVLLPSSLOAGLVSHRELDSSSGPLSVQSLGNFKVWSYLLNPSYLDYLPOLGLRCQ 234  
180 LMWVLLPSSLOAGLVSHRELDSSSGPLSVQSLGNFKVWSYLLNPSYLDYLPOLGLRCQ 239  
235 SSAGGQGVGFRITLTPESCYLSDGCHPOPLRAEMSATAMRRCPPLSTEGLPETHHRRMR 294  
240 NSDGNSEVVGFTLTPESCLRDHCHPOPLSABELIPASWQGCPLSTEGLPETHHRRMR 299  
295 WLVL-OPNQODLPTLDQNGYHSLREHNLRLMDPOHCTDNPAQAVSPAADRP---E 349  
300 RLEFLQAKGQDLPTPDQNGYHSLREHNLRLMDPKHCRDNPQFVPAAGDIPGNTQE 359  
350 PTEKKPELVIOEV-----SQSPQSSSLFCELPVEKECEEDHTNATDLSDRGESLPVSTRP 404  
360 STEEKIELLTTEVPLALEEESPGSPSEIPEKEPGEGRISVVDYLYLEGDLPISARP 419  
405 VCSNKLIDYILGGAPSDLEASDSESDGWEEDDGFSDGSLSDSDVEQDSEGLHLWN 464  
420 ACSNKLIDYILGGASDLETSDDPEGEDWDEAEDDGFSDSLSLSDLEQDPEGLHLWN 479  
465 SFHSVDPYKPNFTATQTAARIAPRDPDSGTSMWSSCGV-GSCQEGPLPETPDHSSGE 523  
480 SFCSDVPYKPNFTATQTAARIAPRDPDSGTSMWSSCGV-GSCQEGPLPETPDHSSGE 539  
524 EDDWEPDAEAEENLKLWNSFCHSDPYNLLNFKAPFQPSGKWKGRQDSKASSEVTVAFS 583  
540 EDDWESSADEAESLKLWNSFCNSDDPYNPLNFKAPFQTSGENEKGCRDSTPSESIVALS 599  
584 GHHTLLSCKAQLLESQDNCPCGGLGALAGERYTHIKRKKVTFLEEVTEYIISGDEDRK 643  
600 ECHTLLSCKVQLLSQSECPDSVQRDVLSGGRHVKRKKVTFLEEVTEYIISGDEDRK 659  
644 GPWEEFARDGCRFQKRIQETEVATGYCLAFEHREKMFNRLR 684  
660 GPWEEFARDGCRFQKRIQETEDAIGYCLTFEHRERMFNRLQ 700

RESULT 4  
US-60-452-680-17655  
; Sequence 17655, Application US/60452680  
; GENERAL INFORMATION:  
; APPLICANT: CARGILL, Michele  
; APPLICANT: GRUPE, Andrew  
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH  
; TITLE OF INVENTION: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001450  
; CURRENT APPLICATION NUMBER: US/60/452,680  
; CURRENT FILING DATE: 2003-03-07  
; NUMBER OF SEQ ID NOS: 116213  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 17655  
; LENGTH: 713  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-60-452-680-17655

Query Match 59.8%; Score 2218; DB 37; Length 713;  
Best Local Similarity 64.9%; Pred. No. 2.1e-183;  
Matches 455; Conservative 54; Mismatches 174; Indels 18; Gaps 7;  
Qy 1 METGTHRRARCPGRLGFWRLPFL-RRSHACSEPPPPSSRQPNGN-----SALPERRT 54  
Db 1 MEPTGGSKRKLGPRAGRFPPPPRRSQSGSKFPTPLGPENSGNPTLLSSAQPETRV 60  
Qy 55 RYWKLLSOLLALPFLFQKLLWSQLSGGLIPTRWLDFAASYSALRASRGRESADPTV 114  
Db 61 SYWKLLSOLLALPGLLQKVLWSQLFGGFPTRWLDFAGVYSALRALKRKPAPTA 120  
Qy 115 QKSLSYTAAGLFAKTRVVTIALARGGTPVAVLRLVLEVKLKAQERALDSAAFTLLEQQ 174  
Db 121 QKSLSSQLD-SSDPSVTSPLDWEELHGWQYSPDDLKLELKAKGSLDPAQAFLLEQQ 179  
Qy 175 LMWVLLPSSLOAGLVSHRELDSSSGPLSVQSLGNFKVWSYLLNPSYLDYLPOLGLRCQ 234

175 LMWVLLPSSLOAGLVSHRELDSSSGPLSVQSLGNFKVWSYLLNPSYLDYLPOLGLRCQ 234  
180 LMWVLLPSSLOAGLVSHRELDSSSGPLSVQSLGNFKVWSYLLNPSYLDYLPOLGLRCQ 239  
235 SSAGGQGVGFRITLTPESCYLSDGCHPOPLRAEMSATAMRRCPPLSTEGLPETHHRRMR 294  
240 NSDGNSEVVGFTLTPESCLRDHCHPOPLSABELIPASWQGCPLSTEGLPETHHRRMR 299  
295 WLVL-OPNQODLPTLDQNGYHSLREHNLRLMDPOHCTDNPAQAVSPAADRP---E 349  
300 RLEFLQAKGQDLPTPDQNGYHSLREHNLRLMDPKHCRDNPQFVPAAGDIPGNTQE 359  
350 PTEKKPELVIOEV-----SQSPQSSSLFCELPVEKECEEDHTNATDLSDRGESLPVSTRP 404  
360 STEEKIELLTTEVPLALEEESPGSPSEIPEKEPGEGRISVVDYLYLEGDLPISARP 419  
405 VCSNKLIDYILGGAPSDLEASDSESDGWEEDDGFSDGSLSDSDVEQDSEGLHLWN 464  
420 ACSNKLIDYILGGASDLETSDDPEGEDWDEAEDDGFSDSLSLSDLEQDPEGLHLWN 479  
465 SFHSVDPYKPNFTATQTAARIAPRDPDSGTSMWSSCGV-GSCQEGPLPETPDHSSGE 523  
480 SFCSDVPYKPNFTATQTAARIAPRDPDSGTSMWSSCGV-GSCQEGPLPETPDHSSGE 539  
524 EDDWEPDAEAEENLKLWNSFCHSDPYNLLNFKAPFQPSGKWKGRQDSKASSEVTVAFS 583  
540 EDDWESSADEAESLKLWNSFCNSDDPYNPLNFKAPFQTSGENEKGCRDSTPSESIVALS 599  
584 GHHTLLSCKAQLLESQDNCPCGGLGALAGERYTHIKRKKVTFLEEVTEYIISGDEDRK 643  
600 ECHTLLSCKVQLLSQSECPDSVQRDVLSGGRHVKRKKVTFLEEVTEYIISGDEDRK 659  
644 GPWEEFARDGCRFQKRIQETEVATGYCLAFEHREKMFNRLR 684  
660 GPWEEFARDGCRFQKRIQETEDAIGYCLTFEHRERMFNRLQ 700

RESULT 3  
US-10-170-2058-23916  
; Sequence 23916, Application US/10170205E  
; GENERAL INFORMATION:  
; APPLICANT: ADAMS, Mark  
; TITLE OF INVENTION: DEVICES, SUCH AS ARRAYS, COMPRISED OF HUMAN PROTEINS OR PROTEIN  
; TITLE OF INVENTION: CAPTURE AGENTS, AND USES THEREOF  
; FILE REFERENCE: CL001381  
; CURRENT APPLICATION NUMBER: US/10/170,205E  
; CURRENT FILING DATE: 2002-06-13  
; NUMBER OF SEQ ID NOS: 40312  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 23916  
; LENGTH: 713  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-170-2058-23916

Query Match 59.8%; Score 2218; DB 27; Length 713;  
Best Local Similarity 64.9%; Pred. No. 2.1e-183;  
Matches 455; Conservative 54; Mismatches 174; Indels 18; Gaps 7;  
Qy 1 METGTHRRARCPGRLGFWRLPFL-RRSHACSEPPPPSSRQPNGN-----SALPERRT 54  
Db 1 MEPTGGSKRKLGPRAGRFPPPPRRSQSGSKFPTPLGPENSGNPTLLSSAQPETRV 60  
Qy 55 RYWKLLSOLLALPFLFQKLLWSQLSGGLIPTRWLDFAASYSALRASRGRESADPTV 114  
Db 61 SYWKLLSOLLALPGLLQKVLWSQLFGGFPTRWLDFAGVYSALRALKRKPAPTA 120  
Qy 115 QKSLSYTAAGLFAKTRVVTIALARGGTPVAVLRLVLEVKLKAQERALDSAAFTLLEQQ 174  
Db 121 QKSLSSQLD-SSDPSVTSPLDWEELHGWQYSPDDLKLELKAKGSLDPAQAFLLEQQ 179  
Qy 175 LMWVLLPSSLOAGLVSHRELDSSSGPLSVQSLGNFKVWSYLLNPSYLDYLPOLGLRCQ 234

180	LKGV	LLPSS	LSQ	RS	L	YS	NR	EL	GSS	PS	G	P	L	Q	R	T	Q	R	D	N	F	S	V	S	V	Y	L	N	P	S	I	L	C	D	F	R	L	E	V	S	Q	239																	
Qy	235	SSAGG	QFQ	VP	RTL	P	ES	C	Y	L	S	E	D	C	H	Q	P	L	A	E	M	S	A	T	A	M	R	C	P	L	S	T	E	G	L	P	E	I	H	R	M	294																	
Db	240	NSDGN	SEVV	GF	Q	T	L	P	ES	S	C	L	R	E	D	H	C	H	Q	P	L	S	A	E	L	I	P	A	S	M	Q	C	P	L	S	T	E	G	L	P	E	I	H	L	R	M	299												
Qy	295	WL	V	L	-	O	P	N	O	G	O	L	P	T	L	D	O	D	N	G	Y	H	S	L	E	E	H	N	L	I	P	M	D	P	Q	H	C	T	D	N	P	A	O	A	V	S	P	A	A	R	P	349							
Db	300	R	L	E	F	Q	A	S	K	G	O	L	P	T	P	O	D	D	N	G	Y	H	S	L	E	E	H	S	L	L	R	M	D	P	K	H	R	D	N	P	T	Q	V	P	A	A	G	I	P	N	T	O	E	359					
Qy	350	P	T	E	K	P	E	L	V	I	Q	B	V	-	-	-	-	S	Q	S	P	S	L	C	F	E	L	P	V	E	K	E	E	D	H	T	N	A	T	L	S	D	R	G	E	L	P	V	S	T	P	404							
Db	360	S	T	E	E	K	I	E	L	L	T	T	E	V	P	L	A	E	E	S	P	E	C	P	S	S	E	I	P	M	E	K	E	G	R	I	S	V	D	V	Y	S	L	E	G	D	L	P	I	S	A	R	P	419					
Qy	405	V	C	S	N	K	L	I	D	I	Y	L	G	A	P	S	D	L	E	A	S	D	S	E	S	H	D	G	E	P	E	D	D	G	F	D	S	D	G	S	L	S	E	S	D	V	Q	D	S	E	G	L	H	L	M	N	464		
Db	420	A	C	S	N	K	L	I	D	I	Y	L	G	A	S	D	L	E	T	S	D	P	E	G	E	D	W	E	A	E	D	D	G	F	D	S	D	S	L	S	D	S	L	E	S	D	L	E	Q	D	P	E	G	L	H	L	M	N	479
Qy	465	S	F	H	S	V	D	P	Y	P	N	F	Q	N	T	A	T	I	O	A	R	I	A	R	P	D	S	D	G	T	S	W	S	G	S	C	G	V	-	G	S	C	Q	G	P	L	P	E	T	P	D	H	S	S	G	E	523		
Db	480	S	F	C	S	V	D	P	Y	P	N	F	Q	N	T	A	T	I	O	A	R	I	V	P	E	E	S	D	E	K	D	L	S	G	S	D	L	E	N	S	Q	G	S	L	P	E	T	P	E	H	S	S	G	E	539				
Qy	524	E	D	D	W	E	S	A	D	E	A	N	L	K	M	N	S	P	C	H	S	E	D	P	Y	N	L	N	F	K	A	P	P	O	P	S	G	N	K	W	K	G	R	O	D	S	K	A	S	E	V	T	V	A	F	S	583		
Db	540	E	D	D	W	E	S	A	D	E	A	S	L	K	M	N	S	F	C	N	S	D	P	Y	N	L	N	F	K	A	P	P	O	T	S	G	E	N	K	G	C	R	D	S	K	T	P	S	E	S	I	V	A	I	S	599			
Qy	584	G	H	T	L	S	C	K	A	O	L	L	E	S	O	E	N	C	P	G	C	G	L	E	A	L	A	C	E	R	Y	T	H	K	K	V	F	L	E	B	T	V	E	Y	I	S	G	D	E	D	R	K	643						
Db	600	E	C	H	T	L	S	C	K	Q	L	G	S	E	S	C	P	D	S	V																																							

## RESULT 5

```

US-09-629-469A-18965
; Sequence 18965, Application US/09629469A
; GENERAL INFORMATION:
; APPLICANT: OTA, TOSHIO
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: NISHIKAWA, TETSUO
; APPLICANT: HAYASHI, KOJI
; APPLICANT: SAITO, KAORU
; APPLICANT: YAMAMOTO, JUNICHI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: WAKAMATSU, AI
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: OTSUKI, TETSUJI
; TITLE OF INVENTION: PRIMERS FOR SYNTHESIZING FULL-LENGTH CDNA AND THEIR USE
; FILE REFERENCE: 084335/0123
; CURRENT APPLICATION NUMBER: US/09/629,469A
; CURRENT FILING DATE: 2000-07-28
; PRIOR APPLICATION NUMBER: JP 1999-248036
; PRIOR FILING DATE: 1999-07-29
; PRIOR APPLICATION NUMBER: JP 1999-300253
; PRIOR FILING DATE: 1999-08-27
; PRIOR APPLICATION NUMBER: JP 2000-118776
; PRIOR FILING DATE: 2000-01-11
; PRIOR APPLICATION NUMBER: JP 2000-183767
; PRIOR FILING DATE: 2000-05-02
; PRIOR APPLICATION NUMBER: JP 2000-241899
; PRIOR FILING DATE: 2000-06-09
; PRIOR APPLICATION NUMBER: 60/159,590
; PRIOR FILING DATE: 1999-10-18
; PRIOR APPLICATION NUMBER: 60/183,322
; PRIOR FILING DATE: 2000-02-17
; NUMBER OF SEQ ID NOS: 19025
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 18965
; LENGTH: 713
; TYPE: PRT

```

```

; ORGANISM: Homo sapiens
us-09-629-469A-18965

Query Match      59.6%; Score 2213; DB 20; Length 7113;
Best Local Similarity 64.8%; Pred. No. 5.8e-183;
Matches 454; Conservative 55; Mismatches 174; Indels 18; Gaps 7;

Qy 1 METGTHRARXRPGRPLGSMFRLPFL-RRSHACSESEPPPSRRQNGN-----SALPERRT 54
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 55 RYTKLLSOLLALLPSLFOKLLWSQLSGGLITRWMLDPAASYALRASHRGRESADPTV 114
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 61 SYTKLLSOLLAPLPGILQKLVLSQSLFGGMFFTRWLDFAGVYSALRALAGREKPAAPTA 120
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 115 QKSLSYTAAGLFAKTRVSTLALARGTTPVAVLVLEVKLKAQERALDSAAFTLEEQ 174
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 121 QKSLSLQLD-SDSPSVTSPLDWLEIGHQWYSPDLKLEKAKGALDDPAQAFLLEEQ 179
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 175 LMGVELLPSSLQAGLVSHRELDSSSGPLASVQSLGNFKVSYLLNPSYLDLPLQLGLRCQ 234
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 180 LMGVELLPSSLQSLRYSLNRELSSGSPGLNIQRIIDFVSVYLLNPSYLDLPCFPRLEVSQ 239
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 235 SSAGGGGVGFRLLTPESCVLSBGCHQPLRAEMSGATAWRRCPPLSTEGLPPIHURMR 294
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 240 NSDGNSEWVGFTLLTPESSCLREDHCHQPLSIAELIPASWQGCPPPLSTEGLPPIHLRMK 299
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 295 WLVPFL-QPNQGDPLPTLDODNGYHSLSEEHNLRLMDPOHCTDNPAQAVSPAADRP---E 349
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 300 RLBFLOQASKGQDLPTPDQDNGYHSLSEEHSLRLMDPKKCRDNPOTFVPAAGDIPGNTOE 359
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 350 PTEKKKPELVIOEV-----SOSPOGSLFCPLPVEKECEEDHTNATDLSRGESLPVSTRP 404
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 360 STEKELELLTTEVPLALEEESPCEGCPSESHIPKEKPGEGRIISVDYSYLEGDLPISARP 419
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 405 VCSNKLIIDYILGAPSDLEASSDSEDMGEPEPDGFDSDGSLSDVQDSEGLHLWN 464
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 420 ACSNKLIIDYILGASSDLETSSPEGEDWDEBAEADDGFDSDSLSDLEQDPEGLHLWN 479
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 465 SFHSVPDYPKQNTATIQTAARAPRDPDSGTSGWSGCV-CSCQEGPLPETHPDHSSGE 523
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 480 SFCVSDPYNPNQNTATIQTAARIIVPEPSSDXDLGKSDLENSSSGSLPETHSSSGE 539
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 524 EDDWESADAENLKLWNSFCHSEDPNLLNPKAPQPSGKNWKGRQDSKASSEVTVAPS 583
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 540 EDDWESADAENSLKLWNSFCNSDDPNPLNFKAPQTSGENEKGCRDSKTPSESIVALS 599
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 584 GHHTLLSCKAQLLESQEDNCPGGLGEALAGERYTHIKRKVTFLBEVTYYISGDEDRK 643
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 600 ECHTLLSCKVQLLGSQSESECPDVSQRDVLSSGRHHTVKKKVTFLFEVTEYYISGDEDRK 659
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 644 GPWEEFARDGCRFQKRIQETVAIGYCLAFEPHEKMFNRLR 684
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 660 GPWEEFARDGCRFQKRIQETDAIGYCLTFEHRERMFNLQ 700

```

## RESULT 6

US-10-917-503-18965  
; Sequence 18965, Application US/10917503  
; GENERAL INFORMATION:  
; APPLICANT: OTA, TOSHIO  
; APPLICANT: ISOGAI, TAKAO  
; APPLICANT: NISHIKAWA, TETSUO  
; APPLICANT: HAYASHI, KOJI  
; APPLICANT: SAITO, KAORU  
; APPLICANT: YAMAMOTO, JUNICHI  
; APPLICANT: ISHII, SHIZUKO  
; APPLICANT: SUGIYAMA, TOMOYASU  
; APPLICANT: WAKAMATSU, AI  
; APPLICANT: NAGAI, KEIICHI  
; APPLICANT: OTSUKI, TETSUJI  
; TITLE OF INVENTION: PRIMERS FOR SYNTHESIZING FULL-LENGTH CDNA AND THEIR USE  
; FILE REFERENCE: 084335/0123

; CURRENT APPLICATION NUMBER: US/10/917,503  
 ; CURRENT FILING DATE: 2004-08-13  
 ; PRIOR APPLICATION NUMBER: US/09/629,469  
 ; PRIOR FILING DATE: 2000-07-28  
 ; PRIOR APPLICATION NUMBER: JP 1999-248036  
 ; PRIOR FILING DATE: 1999-07-29  
 ; PRIOR APPLICATION NUMBER: JP 1999-300253  
 ; PRIOR FILING DATE: 1999-08-27  
 ; PRIOR APPLICATION NUMBER: JP 2000-118776  
 ; PRIOR FILING DATE: 2000-01-11  
 ; PRIOR APPLICATION NUMBER: JP 2000-183767  
 ; PRIOR FILING DATE: 2000-05-02  
 ; PRIOR APPLICATION NUMBER: JP 2000-241899  
 ; PRIOR FILING DATE: 2000-06-09  
 ; PRIOR APPLICATION NUMBER: 60/159,590  
 ; PRIOR FILING DATE: 1999-10-18  
 ; PRIOR APPLICATION NUMBER: 60/183,322  
 ; PRIOR FILING DATE: 2000-02-17  
 ; NUMBER OF SEQ ID NOS: 19025  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 18965  
 ; LENGTH: 713  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; ORGANISM: Homo sapiens  
 US-10-917-503-18965

Query Match 59.6%; Score 2213; DB 35; Length 713;  
 Best Local Similarity 64.8%; Pred. No. 5.8e-183;  
 Matches 454; Conservative 55; Mismatches 174; Indels 18; Gaps 7;  
 Qy 1 METGTHRAKRGPRGLGSWFLPFL-RRSHACSEFPFPPSSRQPN- ----SALPERRT 54  
 Db 1 MEPGTGSRKRLGPRAGFRFPFPPRRSQAGSSKFPTPLGPENSGNPTLLSSAQPETRV 60  
 Qy 55 RYWKLLSOLLALPSLFQKLLWSQLSGGLIPTRWLDFAASYSALRASGRESADPTV 114  
 Db 61 SYWKLLSOLLALPSLFQKLLWSQLSGGLIPTRWLDFAAGVYSALRALGKREKPAAPTA 120  
 Qy 115 QKSLSYTAAGLFAKTRVVTALARGGTPVAVLVRLVLEVKLKAQERALDSAAPTFLLEQQ 174  
 Db 121 QKSLSSQLD-SSDPSTVSPDLWLEEGHGWQSPDLEKLEKAGSALDPAQAFLLEQQ 179  
 Qy 175 LMGVELLPSSLOAGLVSHRELDSSSGPLSVQSLGNFKVSYLLNPYSYLDYLPQLGRCQ 234  
 Db 180 LMGVELLPSSLOAGLVSHRELDSSSGPLSVQSLGNFKVSYLLNPYSYLDYLPQLGRCQ 239  
 Qy 235 SSAGGQFVGFRITLTPESCYLESDGCHPOPLRAEMSAWARRCPPLSTEGLPPIHHRMR 294  
 Db 240 NSDGNSEVVGFTLTPESCCLREDHCHPQLPSAELIPASWQGCPLSTEGLPPIHHRMK 299  
 Qy 295 WLVEL-OPNQGQDLPTLDQDNGYHSLSEEHNLRLMDPOHCTDNPAAQVSPAADRP- -E 349  
 Db 300 RLEFLQAKSQDLPTPDQDNGYHSLSEEHNLRLMDPOHCTDNPAAQVSPAADRP- -E 359  
 Qy 350 PTEKPELVITQEV- - - - -SQSPQSSLFCELPVEKECEDHTNATLSDRGESLPVSTRP 404  
 Db 360 STEEKIELLTTEVPLALEEESPSCGSPSEIPEMEKEGEGRIISVVDVSYLEGDLPIGARP 419  
 Qy 405 VCSNKLIDYILGAPSDLEASSSESDGWEEDDGFDSGSLSESDVDQDSGLHLWN 464  
 Db 420 ACSNKLIDYILGAPSDLEASSSESDGWEEDDGFDSGSLSESDVDQDSGLHLWN 479  
 Qy 465 SFHSVDPYKPNQFTATTQTAARIAPRDPDSGTSWSSGCGV-GSCQSGPLPETPDHSSGE 523  
 Db 480 SFCSVDPYKPNQFTATTQTAARIAPRDPDSGTSWSSGCGV-GSCQSGPLPETPDHSSGE 539  
 Qy 524 EDDWPSADEAENLKLWNSFCHSEDPYNLNFKAFFQPSGKNWGRQDSKASSEVTVAFS 583  
 Db 540 EDDWPSADEAENLKLWNSFCHSEDPYNLNFKAFFQPSGKNWGRQDSKASSEVTVAFS 599  
 Qy 584 GHHTLLSCKAQLLESQEDNCPGCLGALAGERTHYTHIRKKVTFLEEVTEYIISGDEDRK 643  
 Db 600 ECHTLLSCKAQLLESQEDNCPGCLGALAGERTHYTHIRKKVTFLEEVTEYIISGDEDRK 659

Qy 644 GPWEFARDGCRFQKRIQETVAIGYCLAFEPHEKRMFNRLR 684  
 Db 660 GPWEFARDGCRFQKRIQETDAIGYCLTFEPHERMFMRLQ 700  
 RESULT 7  
 US-09-757-028-2348  
 ; Sequence 2348, Application US/09757028  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Rosen et al.  
 ; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
 ; FILE REFERENCE: PM001  
 ; CURRENT APPLICATION NUMBER: US/09/757,028  
 ; CURRENT FILING DATE: 2001-01-09  
 ; PRIOR APPLICATION NUMBER: 60/179,065  
 ; PRIOR FILING DATE: 2000-01-31  
 ; PRIOR APPLICATION NUMBER: 60/180,628  
 ; PRIOR FILING DATE: 2000-02-04  
 ; NUMBER OF SEQ ID NOS: 2660  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 2348  
 ; LENGTH: 720  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; ORGANISM: Homo sapiens  
 US-09-757-028-2348

Query Match 59.5%; Score 2208; DB 22; Length 720;  
 Best Local Similarity 64.8%; Pred. No. 1.6e-182;  
 Matches 454; Conservative 54; Mismatches 175; Indels 18; Gaps 7;  
 Qy 1 METGTHRAKRGPRGLGSWFLPFL-RRSHACSEFPFPPSSRQPN- ----SALPERRT 54  
 Db 8 MEPGTGSRKRLGPRAGFRFPFPPRRSQAGSSKFPTPLGPENSGNPTLLSSAQPETRV 67  
 Qy 55 RYWKLLSOLLALPSLFQKLLWSQLSGGLIPTRWLDFAASYSALRASGRESADPTV 114  
 Db 68 SYWKLLSOLLALPSLFQKLLWSQLSGGLIPTRWLDFAAGVYSALRALGKREKPAAPTA 127  
 Qy 115 QKSLSYTAAGLFAKTRVVTALARGGTPVAVLVRLVLEVKLKAQERALDSAAPTFLLEQQ 174  
 Db 128 QKSLSSQLD-SSDPSTVSPDLWLEEGHGWQSPDLEKLEKAGSALDPAQAFLLEQQ 186  
 Qy 175 LMGVELLPSSLOAGLVSHRELDSSSGPLSVQSLGNFKVSYLLNPYSYLDYLPQLGRCQ 234  
 Db 187 LMGVELLPSSLOAGLVSHRELDSSSGPLSVQSLGNFKVSYLLNPYSYLDYLPQLGRCQ 246  
 Qy 235 SSAGGQFVGFRITLTPESCYLESDGCHPOPLRAEMSAWARRCPPLSTEGLPPIHHRMR 294  
 Db 247 NSDGNSEVVGFTLTPESCCLREDHCHPQLPSAELIPASWQGCPLSTEGLPPIHHRMK 306  
 Qy 295 WLVEL-OPNQGQDLPTLDQDNGYHSLSEEHNLRLMDPOHCTDNPAAQVSPAADRP- -E 349  
 Db 307 RLEFLQAKSQDLPTPDQDNGYHSLSEEHNLRLMDPOHCTDNPAAQVSPAADRP- -E 366  
 Qy 350 PTEKPELVITQEV- - - - -SQSPQSSLFCELPVEKECEDHTNATLSDRGESLPVSTRP 404  
 Db 367 STEEKIELLTTEVPLALEEESPSCGSPSEIPEMEKEGEGRIISVVDVSYLEGDLPIGARP 426  
 Qy 405 VCSNKLIDYILGAPSDLEASSSESDGWEEDDGFDSGSLSESDVDQDSGLHLWN 464  
 Db 427 ACSNKLIDYILGAPSDLEASSSESDGWEEDDGFDSGSLSESDVDQDSGLHLWN 486  
 Qy 465 SFHSVDPYKPNQFTATTQTAARIAPRDPDSGTSWSSGCGV-GSCQSGPLPETPDHSSGE 523  
 Db 487 SFCSVDPYKPNQFTATTQTAARIAPRDPDSGTSWSSGCGV-GSCQSGPLPETPDHSSGE 546  
 Qy 524 EDDWPSADEAENLKLWNSFCHSEDPYNLNFKAFFQPSGKNWGRQDSKASSEVTVAFS 583  
 Db 547 EDDWPSADEAENLKLWNSFCHSEDPYNLNFKAFFQPSGKNWGRQDSKASSEVTVAFS 606  
 Qy 584 GHHTLLSCKAQLLESQEDNCPGCLGALAGERTHYTHIRKKVTFLEEVTEYIISGDEDRK 643









GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 16, 2005, 10:42:43 ; Search time 84.5911 Seconds  
(without alignments)  
1348.829 Million cell updates/sec

Title: US-10-650-482-4

Perfect score: 3710

Sequence: 1 METGTHARKRPGRLGSWF.....MFNRLRIESKDLLLYNVKX 698

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 760079 seqs, 163465437 residues

Total number of hits satisfying chosen parameters: 760079

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

1: Pending\_Patents\_AA New:\*  
2: /cgn2\_6/ptodata/1/paa/PCT\_NEW\_COMB.pep.\*  
3: /cgn2\_6/ptodata/1/paa/US06\_NEW\_COMB.pep.\*  
4: /cgn2\_6/ptodata/1/paa/US07\_NEW\_COMB.pep.\*  
5: /cgn2\_6/ptodata/1/paa/US08\_NEW\_COMB.pep.\*  
6: /cgn2\_6/ptodata/1/paa/US09\_NEW\_COMB.pep.\*  
7: /cgn2\_6/ptodata/1/paa/US10\_NEW\_COMB.pep.\*  
8: /cgn2\_6/ptodata/1/paa/US11\_NEW\_COMB.pep.\*  
9: /cgn2\_6/ptodata/1/paa/US60\_NEW\_COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	189.5	5.1	674	7	US-11-040-219-2
2	175.5	4.7	605	7	US-11-040-219-5
3	151	4.1	3880	1	PCT-US04-37982-2437
4	151	4.1	3880	1	PCT-US04-37982-2439
5	148.5	4.0	676	7	US-11-097-143-13533
6	140.5	3.8	916	6	US-10-899-422-13
7	140.5	3.8	1059	6	US-10-899-422-11
8	139.5	3.8	883	1	PCT-US05-01184A-6
9	138	3.7	913	1	PCT-US04-37982-188
10	138	3.7	913	1	PCT-US04-37982-202
11	138	3.7	1084	1	PCT-US04-37982-192
12	138	3.7	1572	6	US-10-990-328A-11742
13	136.5	3.7	883	1	PCT-US05-01184A-3
14	136.5	3.7	3926	8	US-60-675-841-94
15	136.5	3.7	3926	8	US-60-710-726-316
16	132.5	3.6	1065	6	US-10-940-774A-11618
17	132	3.6	1715	7	US-11-097-143-1059
18	127.5	3.4	3913	6	US-10-940-774A-10933
19	127.5	3.4	4232	8	US-60-651-506-210
20	127.5	3.4	4377	8	PCT-US04-42360-2065
21	127.5	3.4	4377	8	US-60-651-506-211
22	127.5	3.4	4377	8	US-60-651-506-212
23	127	3.4	1028	6	US-10-531-164-756
24	127	3.4	1503	1	PCT-US05-18850-797
25	124.5	3.4	1354	7	US-11-097-143-13869

26	124.5	3.4	1709	1	PCT-US05-21650-40
27	124	3.3	640	6	US-10-450-763-41872
28	124	3.3	1795	6	US-10-450-763-51377
29	123.5	3.3	4274	6	US-10-450-763-31331
30	123.5	3.3	4377	6	US-10-940-774A-6978
31	123.5	3.3	4377	8	US-60-651-506-209
32	123.5	3.3	4386	6	US-10-450-763-37734
33	123.5	3.3	4397	6	US-10-450-763-52303
34	123	3.3	1142	7	US-11-044-051-73
35	123	3.3	1155	6	US-10-450-763-54525
36	123	3.3	1208	6	US-10-450-763-54526
37	122.5	3.3	808	7	US-11-072-512-3401
38	122.5	3.3	961	1	PCT-US03-10870-1618
39	122	3.3	1342	1	PCT-US03-10870-2821
40	121.5	3.3	530	7	US-11-043-591-168
41	121.5	3.3	639	7	US-11-043-591-169
42	121.5	3.3	671	7	US-11-043-591-167
43	120.5	3.2	306	8	US-60-707-841-176
44	120	3.2	1261	6	US-10-940-774A-9651
45	120	3.2	1393	6	US-10-994-106-23
46	120	3.2	1393	6	US-10-994-106-47
47	120	3.2	1393	6	US-10-994-106-58
48	120	3.2	1393	6	US-10-994-106-345
49	120	3.2	1684	1	PCT-US05-00638-84
50	120	3.2	1686	8	US-60-680-002-257
51	120	3.2	1782	1	PCT-US03-10870-2097
52	120	3.2	1784	8	US-60-680-002-256
53	119.5	3.2	1316	6	US-10-450-763-53356
54	119.5	3.2	1321	6	US-10-450-763-55775
55	119	3.2	535	8	US-60-655-875-120587
56	119	3.2	665	7	US-11-043-591-166
57	119	3.2	1040	6	US-10-450-763-51333
58	118.5	3.2	3896	1	PCT-US04-37982-2434
59	118	3.2	1168	6	US-10-990-328A-13049
60	118	3.2	1239	6	US-10-990-328A-13050
61	118	3.2	1239	6	US-10-990-328A-13051
62	118	3.2	1781	8	US-60-680-002-258
63	117.5	3.2	992	7	US-11-090-997-56
64	117.5	3.2	1188	7	US-11-090-997-56
65	117.5	3.2	1413	1	PCT-US04-31416-52
66	117.5	3.2	1423	6	US-10-990-328A-12800
67	117.5	3.2	1698	1	PCT-US04-31416-54
68	117.5	3.2	1709	6	US-10-479-874A-21
69	117.5	3.2	1710	6	US-10-990-328A-12801
70	117.5	3.2	1712	6	US-10-940-774A-9450
71	117.5	3.2	1970	7	US-11-097-143-21273
72	117	3.2	422	7	US-11-072-512-2882
73	117	3.2	780	6	US-10-489-448-3405
74	117	3.2	819	1	PCT-US03-10870-1793
75	117	3.2	1051	7	US-11-097-143-7263
76	117	3.2	1411	8	US-60-710-726-1468
77	116.5	3.1	535	6	US-10-450-763-36563
78	116.5	3.1	535	6	US-10-450-763-44703
79	116.5	3.1	535	6	US-10-450-763-57981
80	116.5	3.1	1454	1	PCT-US03-10870-1328
81	116	3.1	742	6	US-10-538-002-222
82	116	3.1	817	6	US-10-285-830A-8
83	116	3.1	817	7	US-11-041-102-8
84	116	3.1	2312	7	US-11-126-313-34
85	115	3.1	805	6	US-10-940-774A-6713
86	115	3.1	805	7	US-11-037-713-24
87	115	3.1	1468	7	US-11-097-143-15765
88	114.5	3.1	654	6	US-10-781-060-16
89	114.5	3.1	797	6	US-10-204-639-49
90	114.5	3.1	797	6	US-10-204-639A-49
91	114.5	3.1	901	6	US-10-330-773A-870
92	114.5	3.1	1163	7	US-11-097-143-1281
93	114.5	3.1	1210	7	US-11-051-454-264
94	114.5	3.1	1210	8	US-60-710-726-530
95	114	3.1	503	6	US-10-703-032-123990
96	114	3.1	605	5	US-09-833-245B-2237
97	114	3.1	605	6	US-10-934-272-2
98	114	3.1	734	7	US-11-097-143-33975

Sequence 40, Appl  
Sequence 41872, A  
Sequence 51377, A  
Sequence 31331, A  
Sequence 6978, Ap  
Sequence 209, App  
Sequence 37734, A  
Sequence 52303, A  
Sequence 73, Appl  
Sequence 54525, A  
Sequence 54526, A  
Sequence 3401, Ap  
Sequence 1618, Ap  
Sequence 2821, Ap  
Sequence 168, App  
Sequence 169, App  
Sequence 167, App  
Sequence 176, App  
Sequence 9651, Ap  
Sequence 23, Appl  
Sequence 47, Appl  
Sequence 58, Appl  
Sequence 345, App  
Sequence 84, Appl  
Sequence 257, App  
Sequence 2097, Ap  
Sequence 256, App  
Sequence 53356, A  
Sequence 55775, A  
Sequence 120587,  
Sequence 166, App  
Sequence 51233, A  
Sequence 2434, Ap  
Sequence 13049, A  
Sequence 13050, A  
Sequence 13051, A  
Sequence 258, App  
Sequence 59, Appl  
Sequence 56, Appl  
Sequence 52, Appl  
Sequence 12800, A  
Sequence 54, Appl  
Sequence 21, Appl  
Sequence 12801, A  
Sequence 9450, Ap  
Sequence 21273, A  
Sequence 2882, Ap  
Sequence 3405, Ap  
Sequence 1793, Ap  
Sequence 7263, Ap  
Sequence 1468, Ap  
Sequence 36563, A  
Sequence 44703, A  
Sequence 57981, A  
Sequence 1328, Ap  
Sequence 222, App  
Sequence 8, Appl  
Sequence 34, Appl  
Sequence 6713, Ap  
Sequence 24, Appl  
Sequence 15765, A  
Sequence 16, Appl  
Sequence 49, Appl  
Sequence 49, Appl  
Sequence 870, App  
Sequence 1281, Ap  
Sequence 264, App  
Sequence 530, App  
Sequence 123990,  
Sequence 2237, Ap  
Sequence 2, Appl  
Sequence 33975, A



```
Db 486 KVRSEKVTYHFLAVWAGPAQAARQGPWEQLARDRSFARRITQAEBSLCLTPAAR 545
Qy 679 MFNRLR 684
Db 546 AWAHLR 551
```

## RESULT 3

```
PCT-US04-37982-2437
; Sequence 2437, Application PC/TUS0437982
; GENERAL INFORMATION:
; APPLICANT: David W. Morris
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS IN CANCER
; FILE REFERENCE: CHIR0062-500 (23356.0003)
; CURRENT APPLICATION NUMBER: PCT/US04/37982
; CURRENT FILING DATE: 2004-10-21
; PRIOR APPLICATION NUMBER: US 10/692,382
; PRIOR FILING DATE: 2003-10-22
; PRIOR APPLICATION NUMBER: US 10/691,209
; PRIOR FILING DATE: 2003-10-21
; NUMBER OF SEQ ID NOS: 3504
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2437
; LENGTH: 3880
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US04-37982-2437
```

```
Query Match 4.1%; Score 151; DB 1; Length 3880;
Best Local Similarity 22.6%; Pred. No. 3;
Matches 150; Conservative 81; Mismatches 236; Indels 198; Gaps 39;

Qy 9 RKRGP-RLGSWFLPLRLRRSHACSEPPPPSSRQNGNSALPERRTRYWTKLSQLLAL 67
Db 2724 KKKDPLEIGYQAHLPESLSQVSRO-PPKSPQVKPVSPSPHR-----L 2768
Qy 68 LPSLFQKLLWSOLSGGLIPTRMLDFAASYSALRASGRSESDAP--TVQKSL- ---Y 120
Db 2769 LDTSFASLNKAHVS-----PQK--HFTAD--SALR---QOTLPRMKTQRSLSDPKPLS 2817
Qy 121 TAAGLFAKTRVVTALARGTTPVAVLVRLVLEVKLKAQERALSAP-----167
Db 2818 PTAEESAKERF--SLYQHOGGLGSQVSALPNSLVKVKRTLPSPPPEAHLPLAGQDQL 2875
Qy 168 --TFLLEQQLMGVELLPSSLOAGLVSHRELDSSSGPLSVQSLGNFKVSVLLN-----219
Db 2876 YAASSLQRLGTGPTVPAT--KASLL--RELDL-----DLRLVEHSTKLARKKQ 2920
Qy 220 -----PSYLDYLPOLGL--RCOSSA---GGGQFVGFRTLTPESCYLE-----257
Db 2921 AELDEBEKEIDAKLYL--ELGITORKESLAKDRGRDYPPLRGLGEHRDYLSDSELNQLR 2979
Qy 258 -DGC-HPQPLRAEMSATAMRRCPPPLSTEGLEPEIHHRRMRWLVLQP-----NOQDLP 310
Db 2980 LQCTTPAGQFVDFPATA---AAPATPSGPTAQOQR-----FQPPAPQVSAGSGPT- 3029
Qy 311 DDNGYHSLBEEHLLRMDPHQCTDN-----PAQAVSPAADRPEPEKPELVIOE-- 361
Db 3030 --QNGF-----PAHQAPTYPGPSTYPAPAFPPGASYP-----AEPGLPNOQAF 3070
Qy 362 -----VSQSPQ---GSSLFCELPVEKECEDHTNATDLSDRGESLPVSTRPVCSNKLID 412
Db 3071 RPTGHYAGQTPMPTTOSTLF--PVPADSRAPLQKPRQTSLADLEQKVPT-----NYEVID 3123
Qy 413 YI-LGGAPSDLEASSDSESDW--GEEPEDDGFSDGSLSDSESDVEQDSEGLHLWNSFHSV 469
Db 3124 VVPMSSAPSETSYSGPAVSSGVEQKQVEVPRAGDRGSVQSQSPATYPDSH----YTSL 3179
Qy 470 DPYKQPNF-----TATIOAARAPRDPDSGTSWS--GSCGVGSCQEGP-L 513
Db 3180 EQNVPNRYVMIDDISLTAKDSTSTAPDSQRLEPLGSGSRPGKEPBGV---LDGPTL 3236
```

```
Qy 514 PETPDHSSGSEDDWEPFSADAEAKLWNSFCHSEDPYNLLNFKAPPQPSGKNWKGRODSK 573
Db 3237 PCCVREESESDSYDPRGKG-----HLRSMESNGRPASTHYGYDSYR 3280
Qy 574 ASSEV 578
Db 3281 HGARV 3285
```

## RESULT 4

```
PCT-US04-37982-2439
; Sequence 2439, Application PC/TUS0437982
; GENERAL INFORMATION:
; APPLICANT: David W. Morris
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS IN CANCER
; FILE REFERENCE: CHIR0062-500 (23356.0003)
; CURRENT APPLICATION NUMBER: PCT/US04/37982
; CURRENT FILING DATE: 2004-10-21
; PRIOR APPLICATION NUMBER: US 10/692,382
; PRIOR FILING DATE: 2003-10-22
; PRIOR APPLICATION NUMBER: US 10/691,209
; PRIOR FILING DATE: 2003-10-21
; NUMBER OF SEQ ID NOS: 3504
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2439
; LENGTH: 3880
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US04-37982-2439
```

```
Query Match 4.1%; Score 151; DB 1; Length 3880;
Best Local Similarity 22.6%; Pred. No. 3;
Matches 150; Conservative 81; Mismatches 236; Indels 198; Gaps 39;

Qy 9 RKRGP-RLGSWFLPLRLRRSHACSEPPPPSSRQNGNSALPERRTRYWTKLSQLLAL 67
Db 2724 KKKDPLEIGYQAHLPESLSQVSRO-PPKSPQVKPVSPSPHR-----L 2768
Qy 68 LPSLFQKLLWSOLSGGLIPTRMLDFAASYSALRASGRSESDAP--TVQKSL- ---Y 120
Db 2769 LDTSFASLNKAHVS-----PQK--HFTAD--SALR---QOTLPRMKTQRSLSDPKPLS 2817
Qy 121 TAAGLFAKTRVVTALARGTTPVAVLVRLVLEVKLKAQERALSAP-----167
Db 2818 PTAEESAKERF--SLYQHOGGLGSQVSALPNSLVKVKRTLPSPPPEAHLPLAGQDQL 2875
Qy 168 --TFLLEQQLMGVELLPSSLOAGLVSHRELDSSSGPLSVQSLGNFKVSVLLN-----219
Db 2876 YAASSLQRLGTGPTVPAT--KASLL--RELDL-----DLRLVEHSTKLARKKQ 2920
Qy 220 -----PSYLDYLPOLGL--RCOSSA---GGGQFVGFRTLTPESCYLE-----257
Db 2921 AELDEBEKEIDAKLYL--ELGITORKESLAKDRGRDYPPLRGLGEHRDYLSDSELNQLR 2979
Qy 258 -DGC-HPQPLRAEMSATAMRRCPPPLSTEGLEPEIHHRRMRWLVLQP-----NOQDLP 310
Db 2980 LQCTTPAGQFVDFPATA---AAPATPSGPTAQOQR-----FQPPAPQVSAGSGPT- 3029
Qy 311 DDNGYHSLBEEHLLRMDPHQCTDN-----PAQAVSPAADRPEPEKPELVIOE-- 361
Db 3030 --QNGF-----PAHQAPTYPGPSTYPAPAFPPGASYP-----AEPGLPNOQAF 3070
Qy 362 -----VSQSPQ---GSSLFCELPVEKECEDHTNATDLSDRGESLPVSTRPVCSNKLID 412
Db 3071 RPTGHYAGQTPMPTTOSTLF--PVPADSRAPLQKPRQTSLADLEQKVPT-----NYEVID 3123
Qy 413 YI-LGGAPSDLEASSDSESDW--GEEPEDDGFSDGSLSDSESDVEQDSEGLHLWNSFHSV 469
Db 3124 VVPMSSAPSETSYSGPAVSSGVEQKQVEVPRAGDRGSVQSQSPATYPDSH----YTSL 3179
Qy 470 DPYKQPNF-----TATIOAARAPRDPDSGTSWS--GSCGVGSCQEGP-L 513
```



Db 476 DAPPKDVPPCQBPAPPAQDLSPOCDLPAGQEPPLPHQDPLLTOKDLPATQESPTRDLPPCQD 535  
QY 375 -----LPVEKECEEDHTNATDLSRGSLSLVSTRPV-----C 406  
Db 536 LPPSQVSLPA-KALTEDTMSSGDLAATGDPAPAPAFVIVPVRDSTYSQKAGAEQC 594  
QY 407 SNKLIDYILGAPSDLEASDSSESDEWGEPEDDGFDSD--GSLSES-----DVEQDSE 458  
Db 595 S-----GDEBDAEAEVEGEDEDEDTSDNYGERSEAKRSMIETGQAE 645  
QY 459 G-----LHLWNSF-----HSVDYKPNQFTATIQTAAARIAPRP-----SDSGTWSGCGVGC 508  
Db 646 GGLSLRVQNSLRRRTHS-----EGSLLQEPGRCFASDTTLHCSDGEGAAT 692  
QY 509 QEGPLPET-----PDHSS-----GEEDD----- 526  
Db 693 WGMPSPTLKKELGRNGSMHLSLFTGHRKMSGADTVGDDDEASRKRKSKNRAKDMKN 752  
QY 527 -----WEPSSADEA-----ENLKLWNSFCHS 546  
Db 753 KLGIFRRRNEPSPAGPAGKADMMKSPRTSEALKWGESLEKLLVHKDGLAVFQAFRLT 812  
QY 547 EDPYLNLFKAPFOPSGKNW-----KGRDQSKASSEVTVAFSGHHTLLSCQAQLLES- 598  
Db 813 EFSSENLFF-----WLACEDFKVKVQSQSKWAKKIFABYIAIOACKEVNLDYS 862  
QY 599 -----QEDNCP-----GC 606  
Db 863 TREHTKDNLSQSVTRGC 878

RESULT 7  
US-10-899-422-11  
; Sequence 11, Application US/10899422  
; GENERAL INFORMATION:  
; APPLICANT: Lorientis limited  
; TITLE OF INVENTION: Medical Treatment  
; FILE REFERENCE: P020742US CLM  
; CURRENT APPLICATION NUMBER: US/10/899,422  
; CURRENT FILING DATE: 2004-07-26  
; PRIOR APPLICATION NUMBER: GB0201674.9  
; PRIOR FILING DATE: 2002-01-25  
; NUMBER OF SEQ ID NOS: 54  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 11  
; LENGTH: 1059  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-899-422-11

Query Match 3.8%; Score 140.5; DB 6; Length 1059;  
Best Local Similarity 19.2%; Pred. No. 2.4;  
Matches 165; Conservative 82; Mismatches 243; Indels 371; Gaps 37;  
QY 5 THRARKRPGPR-----LGSWFR-----LPFLRRSHACSSSEPPPP 38  
Db 273 THGVQARPEQRHSHLVCDSSDGLLGWERYTEVAKRGQHTLPALSRAATPDNVI 332  
QY 39 SSRONPGNSALPERRTRYWTKLSQLLALLPSLFQKLLWSQLSGGLIPTRWLDFASYS 98  
Db 333 LAPLNPQSQ-----LRPVYQE-----DTIP----- 353  
QY 99 ALRASRGRESDAPTVOKLSYTAAGLFAKTRVYSTLALAG----- 140  
Db 354 -----EEGGSFS-----GKSYT-----GLGKSRMLKVTQTMKGHNYQNCVPRPHATHSS 401  
QY 141 -GTPVAVLVLRLEVKLKAQERALDSAAFTFLLEQQLMGVELLPSSLOAGLVSHREL----- 195  
Db 402 YGTVVTLAPKLVLPVFPVQPLDLCNPARTLLSELLLYEGRNKAEEVTLFAYSDDLFT 461  
QY 196 DSSSGSGLSVQSLGNFKVSVYLLNPSYLDYLPQLGLRQSSAGGQFVGFRTLTPESCYL 255  
Db 462 KEDBFRGCDV-----LRNPLYQSV-----KLQEGSSSDLKFCVL-----YL 498

QY 256 SEDGCHPOPLRAEMSNATAWRR-----PPLSTEGLPETIHRMRMLVFL 299  
Db 499 AEKAECLFTLEAHSQKQKRVCMCLSENIAKQOOLAASPPDSKKLHTDADHKE-----M 553  
QY 300 QPNQ-----GQDL-PTLDODNGYHSEEHNLRLMD----- 329  
Db 554 APDKANKGAEDSPSPKESPGQELPPGQDLPPNKDPSGQEPAPQOEPLSSKDSATSEG 613  
QY 330 -----PQCHTNPQAQVSPAADRP-----BPTKKKPELVIQE-----VSQSPQSS 370  
Db 614 SPPGPDAPSKDVPQCEPPPAQDLSPCQDLPAQEPPLPHQDPLLTOKDLPATQESPTRDL 673  
QY 371 LFCE-----LPVEKECEEDHTNATDLSRGSLSLVSTRPV----- 405  
Db 674 PCQDLPPSQVSLPA-KALTEDTMSSGDLAATGDPAPAPAFVIVPVRDSTYSQKAG 732  
QY 406 -----CSNKLDIYILGAPSDLEASDSSESDEWGEPEDDGFDSD--GSLSES-----DV 453  
Db 733 AEGGCS-----GDEBDAEAEVEGEDEDEDTSDNYGERSEAKRSMIET 783  
QY 454 EQDSEG-----IHLWNSF-----HSVDYKPNQFTATIQTAAARIAPRP-----SDSGTWSGSC 503  
Db 784 GQAEGLSLRVQNSLRRRTHS-----EGSLLQEPGRCFASDTTLHCSDGE 830  
QY 504 GVSGCEGPLPET-----PDHSS-----GEEDD----- 526  
Db 831 GAASTWGMPSPTLKKELGRNGSMHLSLFTGHRKMSGADTVGDDDEASRKRKSKNRA 890  
QY 527 -----WEPSSADEA-----ENLKLWNSFCHS 541  
Db 891 KDMKNKLGIFRRRNEPSPAGPAGKADMMKSPRTSEALKWGESLEKLLVHKDGLAVFQ 950  
QY 542 SFCHSEDPYLNLFKAPFOPSGKNW-----KGRDQSKASSEVTVAFSGHHTLLSCQA 594  
Db 951 AFLRTFSENLFF-----WLACEDFKVKVQSQSKWAKKIFABYIAIOACKEV 1000  
QY 595 LLES-----QEDNCP-----GC 606  
Db 1001 NLDSYTREHTKDNLSQSVTRGC 1021

RESULT 8  
PCT-US05-01184A-6  
; Sequence 6, Application PC/TUS0501184A  
; GENERAL INFORMATION:  
; APPLICANT: Yale University  
; APPLICANT: Hockfield, Susan  
; APPLICANT: Matthews, Russell T.  
; APPLICANT: Viapiano, Mariano S.  
; TITLE OF INVENTION: PRIMARY CENTRAL NERVOUS SYSTEM TUMOR SPECIFIC BEHAV ISOFORMS  
; FILE REFERENCE: 044574-5126-WO  
; CURRENT APPLICATION NUMBER: PCT/US05/01184A  
; CURRENT FILING DATE: 2005-01-14  
; PRIOR APPLICATION NUMBER: US 60/536,594  
; PRIOR FILING DATE: 2004-01-15  
; NUMBER OF SEQ ID NOS: 9  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 6  
; LENGTH: 883  
; TYPE: PRT  
; ORGANISM: Rattus rattus  
PCT-US05-01184A-6

Query Match 3.8%; Score 139.5; DB 1; Length 883;  
Best Local Similarity 22.3%; Pred. No. 2.2;  
Matches 154; Conservative 66; Mismatches 217; Indels 253; Gaps 39;  
QY 140 GGTPTVAVLV---LRLEVKLKAQERALDSAAFTFLLEQQLMGVELLP---SSLOAGLVSHRE 194  
Db 83 GDREVEVLVARGVRKVNENYRFRVALPAYPASLTDVSLVLSRLPNDSGVYRCEVQHG 142  
QY 195 LDSSSGSGLSVQSL-----GNPKVSVYLLNPSYLDYLPQLGLRQ 234

```
Db 143 DDSSDAVEKVKGVFLYREGSARYAFSFAQAQACARIGARIATPEQL-VAAYLGGVEQ 201
Qy 235 SSAG--GGQVGFRTLT-ESCYLSEGG-----CHPQLRAEMSA- 271
Db 202 CDAGWLSQDTVRIPIQNPACACGDMGYPGVNRYGVVDPDLVDVYCYAEDLNGELFLG 261
Qy 272 -----TAWR---RCP-----PLSTE----- 283
Db 262 APPGKLTWEARDYCLERGAQIASTGQLYAAWNGGLDCSPGWLADGVSRYPIITPSQRC 321
Qy 284 --GLPEIHRMRWLVLFPQNGQDLPTLDQDNGVHSLBEEHNLRLMDPQCTONPA--- 338
Db 322 GGGLPGVK-----TLFLFPNQ-----TGPPSKQNRNV-----YCFRDSAHPS 359
Qy 339 ---QAVSPAARDPEP---TKKPELVI-QEVSQSPQSGSLFCELPVKECEEDHTNATD 390
Db 360 AFSEASSPASDGLAIVTVTKLELQLPQBAVSESGAIY-SIPI-----TE 407
Qy 391 LSDRGESLPVSTRPVCSNKLIDYILGGAP-SDLEASSDS-----ESDMDGEEPEDDGFD 443
Db 408 DGGGGSSTPEP-----AEAPRTLESETQSVAPPTGCSSEEGEALBEEERF 454
Qy 444 SDGSLSESDVQDSGLHLW-----NSFHSVDPYKPNFTATIQTAAARIAPRD 491
Db 455 KDTETPKKEQ--ENLWVWPTLSSPLTGLETHSLSQVSP-QAVALQLGASPPRP 511
Qy 492 PSDGTSWSGCGVSCOGGLPETPDHS---SGEEDDWEPSADEAENLKLWNSFCHSED 548
Db 512 PRVHGPP---AETQPPREGSLTSPDGAREVAGETGSPFELSGVPRESEEGSS--SLED 566
Qy 549 PYNLINFKAQPPQSG-----KNWKR-----QDSKASSEVTVAFSGH 585
Db 567 GPSLL--PATWAPVGTRELTPESEKSGRTVLTGTSVQAQVLPVTDASARGVAVAPSSG 624
Qy 586 HTLIS-CK--AQLLESQD-NC--PG-----CGLG-----BALAGERYTHIKRKV 625
Db 625 DCIPSPCHNGGTCLKEEGFRCLPLPGYGGDLCDVGLHFCSPGWEAFQACYKHFSTR-- 683
Qy 626 TFLBEVTEYISGDEDRKGPWEFARDGCR 655
Db 684 -----SWE-EAESQCR 693

RESULT 9
PCT-US04-37982-188
; Sequence 188, Application PC/TUS0437982
; GENERAL INFORMATION:
; APPLICANT: David W. Morris
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS IN CANCER
; FILE REFERENCE: CHIR0062-500 (23356.0003)
; CURRENT APPLICATION NUMBER: PCT/US04/37982
; PRIOR FILING DATE: 2004-10-21
; PRIOR APPLICATION NUMBER: US 10/692,382
; PRIOR FILING DATE: 2003-10-22
; PRIOR APPLICATION NUMBER: US 10/691,209
; PRIOR FILING DATE: 2003-10-21
; NUMBER OF SEQ ID NOS: 3504
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 188
; LENGTH: 913
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US04-37982-188

Query Match 3.7%; Score 138; DB 1; Length 913;
Best Local Similarity 19.9%; Pred. No. 2.8;
Matches 164; Conservative 84; Mismatches 266; Indels 310; Gaps 34;

Qy 5 THRAKRKPPR-----LGSWFR-----LPFLRRSHACSSERPPP 38
Db 128 THGVOARPEQRHCHLVCDSSGILLGGWERYTEVAKRGQGHQTLPAUSRATAPTDPNYII 187
```

```
Qy 39 SSRQNPQNSALPERRTRYTKLISQLLALLPSLFQKLLWSQLSGGIIPTRWLDDFAASYS 98
Db 188 LAPLPGSQ-----LLRPVQE-----DPIP----- 208
Qy 99 ALRASRGRESDPVTKVKSLSYTAAGLFAKTRVVSTLALARG----- 140
Db 209 -----EBSGSPS--KGKSYT--GLGKESRLMKTVQTMKGHGNQYQCPVVRPHATHSS 256
Qy 141 -GTPVAVLVRLEVKLKAQERALDSAAPTFLEQOLMGVELLPSLSQAGLVSHREL--- 195
Db 257 YGTVTVLAPKVLVPFVQDPLDLCNPARTLLSEELLLYEGRNKAAEVTILPAYSDLLFT 316
Qy 196 DSSSGSPLSVQSLGNFKVSVYLLNPSYLDLPQLGLRCQSSAGGQGVGFRTILTPESCYL 255
Db 317 KEDEPGRCDV-----LRNPLYQSV-----KLQEGSSEDLKFCVL-----YL 353
Qy 256 SEDG-----CHPQPIR-----AEMSATAMRRCPLSTEGPLPEIHRMRWL----- 296
Db 354 AEKAECLFTLEAHSQEQKRVCMCLSENIAKQOQLADPDSKMFETEADKREMALEBKG 413
Qy 297 -----VFLQPNQDQL-----PTLDQDNGVHSLBEEHNLRLMD----- 329
Db 414 PGABDSPSPSKEPSPGQDLPPGQDLPPNKSQSPQEPAPSQBPSSKDSATSESGPPGPD 473
Qy 330 -----PQHCTDPAQAVSPAARDP---EPTKKPELVIQE---VSQSPQSGSLFCE--- 374
Db 474 PPSKDVPPCOEPPPAQDLSPQDLPAQGEPLPHQDPLLTOKDLPAIQESPTRDLPQCQLP 533
Qy 375 -----LPVEKECEDHTNATLSDRGESLSPVSTRPV-----CSN 408
Db 534 PSQVSLPA-KALTEDTMSSGDLAAATGDPAPAPPAFVPIEVRIDSTYSQKAGAEQCS- 591
Qy 409 KLIDYILGGAPSDLEASSDSESDMGPEPDDGPDSD--GSLSES-----DVEQDSEG- 459
Db 592 -----GDEEABAEABEVEEGEEDDEDSDNYGERSEAKSSMIETCGGEGG 643
Qy 460 --LHLWNSF---HSDVPYKPNFTATIQTAAARIAPRD---SDSGTSWSGSCGVGSCQE 510
Db 644 LSLRVQNSLRRRTHS-----EGSLLQEPGRCFASDPTLHCSDEGEGAASTWG 690
Qy 511 GPLPETPDHSSGE-----EDDWEPSADEAENLKLWNSFCH 545
Db 691 MPSPFTLLKELGRNGSGMHHLSLFTGHRKQSGADTVGDDDEASRKKKXNL----- 742
Qy 546 SEDPYNLL-----NFKAPFQPGKQWKGRQDSKASSEVTVAFSGHHTLLSCKAQLLESQ 599
Db 743 AKQWKNKLGIFRRRNESPGAPPAGKADKMKSPKPTSEALKW-----GESLEKL 792
Qy 600 EDNCPGCGLGEALAGERYTHIKRKVKVTFLEBVEYIISGDEDRK 643
Db 793 LVHKYGLAVFQAFL-----RTEFSEENLEFWLACEDFKK 826

RESULT 10
PCT-US04-37982-202
; Sequence 202, Application PC/TUS0437982
; GENERAL INFORMATION:
; APPLICANT: David W. Morris
; APPLICANT: Marc S. Malandro
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS IN CANCER
; FILE REFERENCE: CHIR0062-500 (23356.0003)
; CURRENT APPLICATION NUMBER: PCT/US04/37982
; CURRENT FILING DATE: 2004-10-21
; PRIOR APPLICATION NUMBER: US 10/692,382
; PRIOR FILING DATE: 2003-10-22
; PRIOR APPLICATION NUMBER: US 10/691,209
; PRIOR FILING DATE: 2003-10-21
; NUMBER OF SEQ ID NOS: 3504
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 202
; LENGTH: 913
; TYPE: PRT
```

```
; ORGANISM: Homo sapiens
PCT-US04-37982-202

Query Match
  3.7%; Score 138; DB 1; Length 913;
Best Local Similarity 19.9%; Pred. No. 2.8;
Matches 164; Conservative 84; Mismatches 266; Indels 310; Gaps 34;

Qy 5 THRARKRPGPR-----LGSWFR-----LPFLRRSHACSESEPPPP 38
Db 128 THGVQARPEQRHSHLVCDSSDGLLGGWERYTEVAKRGQHTLPALSRATAPDNYII 187
Qy 39 SSRQNPNSALPERTRYWTKLLSQALLPSLFQKLLWSQLSGGLIPTRWLDFAASYS 98
Db 188 LAPLPGSQ-----LRLPVYQE-----DTIP----- 208
Qy 99 ALRASRGRESDAPTQKSLSYTAAGLFAKTRVVSTLALARG----- 140
Db 209 -----BESGSPS--KGKSYT--GLGKKSLRMKTQTMKGHNQNCVVRPHATHSS 256
Qy 141 -GTPVAVLVLRLEVKLKAQERALDSAPTFLLEQOLWGVELLPSSLOAGLVSHREL----- 195
Db 257 YGTVTTLAPKLVFPVFPVQPLDLCNPARTLLSBEILLYEGRNKAAEVTLPAYSDLLFT 316
Qy 196 DSSSGPLSVQSLGNFKVSYLLNPSYLDYLPOLGLRCQSSAGGQFVGRTLTPESCYL 255
Db 317 KEDEPGRCDV-----LRNPLYLQSV-----KLOEGSSEDLKFCVL-----YL 353
Qy 256 SEDG-----CHPQPLR-----AEMSATAWRRCPPLSTEGGLPEIHRRMRWL----- 296
Db 354 AEKAECLFTLEAHSEQKRVKWCCLSENIAKQQLADPDQKMPTEADEKREMALEBEGK 413
Qy 297 -----VFLQPNQGD-----PTLDQNGYHSLSEEHNLRLMD----- 329
Db 414 PGAEDSPSKBPSPGQELPPGQDLPPNKDPSGQEPAPSQEPPLSSKDSATSESGPPGPD 473
Qy 330 -----POHCTDNPAQAVSPAADRP---EPTKEKPELVQE---VSQSPQSSSLFCE--- 374
Db 474 PPSKDVPPCOEPPPAQDLSPCQDLPAQOEPLPHQDPLLTKDLPALQIESPTRDLPPCQDLP 533
Qy 375 -----LPVKECEDHTNATDLSDRGESLPVSTRPV-----CSN 408
Db 534 PSQVSLPA-KALTEDTWSGDLAATGDPAPAPAFVPEVRLDSTYSQKAGAEQCS- 591
Qy 409 KLIDYILGAPSDLEASDSEDEWGEDDGFDS-GLSLES-----DVEQDSEG- 459
Db 592 -----GDEDAEAEAEVEGEDEDEDTSDNYGERSSEAKRSSMIETCQGAEGG 643
Qy 460 --LHLWNSF-----HSVDVPKPNFTATIQTAARAPRDP---SDSGTSMWSCGCVGSCQE 510
Db 644 LSLRVQNSLRRTTHS-----EGSLLOEPRGCFASDTTLHCSDEGGAATWG 690
Qy 511 GPLPETPDHSSGE-----EDDWPSADAENLKLWNSFCH 545
Db 691 MPSPSTLKKELGRNGGSMHLSLFFTGHRKWSGADTVGDDDEASRKRKSNL----- 742
Qy 546 SEDPYNLL-----NFKAPPQPSGKWKGRQDSKASSEVTVAFGHHTLLSCKAQLLESQ 599
Db 743 AKDMKNKLGIFRRRNESPGAPPAGKADKMKWSFKPTSEALKW-----GESLEKL 792
Qy 600 EDNCPGCGLGEALAGERYTHIKRKKVTFLEEVTEYVIGDEDRK 643
Db 793 LVHKYGLAVFQAPL-----RTEFSEENLEFWLACEDFKK 826
```

RESULT 11

```
PCT-US04-37982-192
; Sequence 192, Application PC/TUS0437982
; GENERAL INFORMATION:
; APPLICANT: David W. Morris
; APPLICANT: Marc S. Malandro
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS IN CANCER
; FILE REFERENCE: CHIR0062-500 (23356.0003)
; CURRENT APPLICATION NUMBER: PCT/US04/37982
```

```
; CURRENT FILING DATE: 2004-10-21
; PRIOR APPLICATION NUMBER: US 10/692,382
; PRIOR FILING DATE: 2003-10-22
; PRIOR APPLICATION NUMBER: US 10/691,209
; PRIOR FILING DATE: 2003-10-21
; NUMBER OF SEQ ID NOS: 3504
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 192
; LENGTH: 1084
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US04-37982-192
```

```
Query Match
  3.7%; Score 138; DB 1; Length 1084;
Best Local Similarity 19.9%; Pred. No. 3.6;
Matches 164; Conservative 84; Mismatches 266; Indels 310; Gaps 34;

Qy 5 THRARKRPGPR-----LGSWFR-----LPFLRRSHACSESEPPPP 38
Db 299 THGVQARPEQRHSHLVCDSSDGLLGGWERYTEVAKRGQHTLPALSRATAPDNYII 358
Qy 39 SSRQNPNSALPERTRYWTKLLSQALLPSLFQKLLWSQLSGGLIPTRWLDFAASYS 98
Db 359 LAPLPGSQ-----LRLPVYQE-----DTIP----- 379
Qy 99 ALRASRGRESDAPTQKSLSYTAAGLFAKTRVVSTLALARG----- 140
Db 380 -----BESGSPS--KGKSYT--GLGKKSLRMKTQTMKGHNQNCVVRPHATHSS 427
Qy 141 -GTPVAVLVLRLEVKLKAQERALDSAPTFLLEQOLWGVELLPSSLOAGLVSHREL----- 195
Db 428 YGTVTTLAPKLVFPVFPVQPLDLCNPARTLLSBEILLYEGRNKAAEVTLPAYSDLLFT 487
Qy 196 DSSSGPLSVQSLGNFKVSYLLNPSYLDYLPOLGLRCQSSAGGQFVGRTLTPESCYL 255
Db 488 KEDEPGRCDV-----LRNPLYLQSV-----KLOEGSSEDLKFCVL-----YL 524
Qy 256 SEDG-----CHPQPLR-----AEMSATAWRRCPPLSTEGGLPEIHRRMRWL----- 296
Db 525 AEKAECLFTLEAHSEQKRVKWCCLSENIAKQQLADPDQKMPTEADEKREMALEBEGK 584
Qy 297 -----VFLQPNQGD-----PTLDQNGYHSLSEEHNLRLMD----- 329
Db 585 PGAEDSPSKBPSPGQELPPGQDLPPNKDPSGQEPAPSQEPPLSSKDSATSESGPPGPD 644
Qy 330 -----POHCTDNPAQAVSPAADRP---EPTKEKPELVQE---VSQSPQSSSLFCE--- 374
Db 645 PPSKDVPPCOEPPPAQDLSPCQDLPAQOEPLPHQDPLLTKDLPALQIESPTRDLPPCQDLP 704
Qy 375 -----LPVKECEDHTNATDLSDRGESLPVSTRPV-----CSN 408
Db 705 PSQVSLPA-KALTEDTWSGDLAATGDPAPAPAFVPEVRLDSTYSQKAGAEQCS- 762
Qy 409 KLIDYILGAPSDLEASDSEDEWGEDDGFDS-GLSLES-----DVEQDSEG- 459
Db 763 -----GDEDAEAEAEVEGEDEDEDTSDNYGERSSEAKRSSMIETCQGAEGG 814
Qy 460 --LHLWNSF-----HSVDVPKPNFTATIQTAARAPRDP---SDSGTSMWSCGCVGSCQE 510
Db 815 LSLRVQNSLRRTTHS-----EGSLLOEPRGCFASDTTLHCSDEGGAATWG 861
Qy 511 GPLPETPDHSSGE-----EDDWPSADAENLKLWNSFCH 545
Db 862 MPSPSTLKKELGRNGGSMHLSLFFTGHRKWSGADTVGDDDEASRKRKSNL----- 913
Qy 546 SEDPYNLL-----NFKAPPQPSGKWKGRQDSKASSEVTVAFGHHTLLSCKAQLLESQ 599
Db 914 AKDMKNKLGIFRRRNESPGAPPAGKADKMKWSFKPTSEALKW-----GESLEKL 963
Qy 600 EDNCPGCGLGEALAGERYTHIKRKKVTFLEEVTEYVIGDEDRK 643
Db 964 LVHKYGLAVFQAPL-----RTEFSEENLEFWLACEDFKK 997
```







**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 16, 2005, 08:56:27 ; Search time 17.8086 Seconds  
(without alignments)  
3771.163 Million cell updates/sec

Title: US-10-650-482-4

Perfect score: 3710

Sequence: 1 METGTHRAKRPGRGLGSWF.....MFNRLRIESKLLLYSNVKK 698

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : PIR 79.\*

1: pir1.\*

2: pir2.\*

3: pir3.\*

4: pir4.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	204	5.5	657	2 S10001	MyD116 protein - m
2	171.5	4.6	590	2 A56535	gadd34 protein - l
3	151.5	4.1	1510	2 T33100	hypothetical prote
4	141	3.8	522	2 T40520	hypothetical prote
5	136.5	3.7	1571	2 T00062	hypothetical prote
6	135	3.6	527	2 C88042	protein P56D12.4 [
7	133.5	3.6	782	2 T32155	hypothetical prote
8	133.5	3.6	883	2 S49136	brevican precursor
9	132.5	3.6	682	2 P90603	vipe-like (mycopla
10	130	3.5	613	2 T42671	hypothetical prote
11	130	3.5	1560	2 T30282	calcium-binding pr
12	129.5	3.5	1473	2 T31422	C-terminal domain-
13	127.5	3.4	406	2 S38170	SRP40 protein - ye
14	126	3.4	817	2 T03852	protein phosphatas
15	126	3.4	1229	2 T46116	hypothetical prote
16	123.5	3.3	4377	2 A55575	ankyrin 3, long sp
17	123	3.3	1324	2 T01508	mismatch repair en
18	123	3.3	4549	2 T20771	hypothetical prote
19	123	3.3	4667	2 T20774	hypothetical prote
20	122	3.3	1189	2 S56852	hypothetical prote
21	121.5	3.3	953	2 T01093	luminidependens pr
22	120.5	3.2	306	1 BVBK6	MAK16 protein - ye
23	120.5	3.2	660	2 JW0037	homeotic protein H
24	120	3.2	1684	2 T01508	gravin - human
25	119.5	3.2	961	2 A55380	faciogenital dyspl
26	119.5	3.2	1316	2 T00381	KIAA0633 protein -
27	119.5	3.2	1325	2 S62497	probable nucleopor
28	119.5	3.2	1468	2 S11515	formin - mouse
29	119.5	3.2	3942	2 T42730	Bassoon protein -

30	118.5	3.2	313	2 G84413	hypothetical prote
31	118	3.2	1026	2 T28968	hypothetical prote
32	118	3.2	1239	2 G02750	DNA-directed DNA p
33	118	3.2	1829	2 T26135	hypothetical prote
34	118	3.2	4342	2 H83343	probable non-ribos
35	117.5	3.2	883	2 S57653	brevican precursor
36	117.5	3.2	925	2 T48834	hypothetical prote
37	117.5	3.2	5762	2 A41819	proline-rich pepti
38	117	3.2	184	2 A42749	hypothetical prote
39	117	3.2	1231	2 S30185	insulin receptor s
40	117	3.2	1317	2 T03748	apoptosis associat
41	117	3.2	1341	2 T17285	hypothetical prote
42	116.5	3.1	912	2 A54423	brevican precursor
43	116.5	3.1	1331	2 AE1843	hypothetical prote
44	116	3.1	699	2 I38073	nucleolar phosphop
45	116	3.1	1300	2 T03166	probable immediate
46	116	3.1	1974	2 T16703	hypothetical prote
47	115.5	3.1	1621	2 A82255	hypothetical prote
48	115	3.1	1241	2 JU0466	potassium transpor
49	114.5	3.1	860	2 T39502	hypothetical prote
50	114.5	3.1	1163	2 A36685	205K microtubule-a
51	114.5	3.1	1210	2 I39410	AF-4 protein, spli
52	114.5	3.1	1213	2 A58198	serine/proline-ric
53	114.5	3.1	1485	1 IS2PT2	DNA topoisomerase
54	114.5	3.1	2524	2 A35844	Xotch protein - Af
55	114	3.1	1463	2 T30290	AAS surface protei
56	114	3.1	2437	2 S53611	MBPI protein - ra
57	113.5	3.1	699	2 AC0504	Probable chitinase
58	113.5	3.1	897	2 E65127	Probable bifunctio
59	113.5	3.1	1180	2 T20773	hypothetical prote
60	113	3.0	419	1 TVLUMC	transforming prote
61	113	3.0	823	2 T21104	hypothetical prote
62	113	3.0	1495	2 T48429	hypothetical prote
63	113	3.0	2009	2 S49764	SEC7 protein - yea
64	112.5	3.0	382	2 S72169	DNA-binding protei
65	112.5	3.0	681	2 P85062	hypothetical prote
66	112.5	3.0	1805	2 A34736	nestin - rat
67	112	3.0	1185	2 A42404	collagen adhesin -
68	112	3.0	1237	2 A34598	ecdysone-induced p
69	112	3.0	1394	2 B34598	ecdysone-induced p
70	112	3.0	3938	2 T42761	Bassoon protein -
71	111.5	3.0	364	1 EDBESP	immediate-early pr
72	111.5	3.0	619	2 T08613	hypothetical prote
73	111.5	3.0	757	2 S68142	probable transcrip
74	111.5	3.0	1249	2 T14150	vesicle associated
75	111.5	3.0	3707	2 S18252	heparan sulfate pr
76	111	3.0	549	2 T40692	hypothetical prote
77	111	3.0	1235	1 S16948	insulin receptor s
78	110.5	3.0	1791	2 T02345	hypothetical prote
79	110.5	3.0	764	2 JCS113	ribosomal transcri
80	110.5	3.0	727	2 JCS112	ribosomal transcri
81	110.5	3.0	906	2 A43817	transforming prote
82	110.5	3.0	930	2 D37271	A-alpha 2 4 protei
83	110.5	3.0	932	2 T48761	hypothetical prote
84	110	3.0	489	2 S15349	mdm2 protein - mou
85	110	3.0	509	2 T34871	probable membrane
86	110	3.0	649	2 T46500	hypothetical prote
87	110	3.0	816	2 T17257	hypothetical prote
88	110	3.0	1487	2 T02850	hypothetical prote
89	110	3.0	2688	2 T49477	alpha-A-crystallin
90	109.5	3.0	842	2 T23715	hypothetical prote
91	109.5	3.0	1062	2 T46444	hypothetical prote
92	109.5	3.0	1093	2 I38533	AF17 protein - hum
93	109.5	3.0	1231	2 T18532	serine/threonine s
94	109.5	3.0	1242	2 JS0670	neulin receptor s
95	109.5	3.0	1286	2 T23714	hypothetical prote
96	109.5	3.0	1596	2 A35927	190K DNA-binding p
97	109.5	3.0	1624	2 T25592	hypothetical prote
98	109.5	3.0	1757	2 T05204	hypothetical prote
99	109.5	3.0	2774	2 A43359	microtubule-associ
100	109.5	3.0	3924	2 S37431	ankyrin 2, neurona

## ALIGNMENTS

## RESULT 1

S10001  
MyD116 protein - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 31-Dec-1990 #sequence\_revision 19-Dec-1990 #text\_change 09-Jul-2004  
C:Accession: S10001  
R:Lord, K.A.; Hoffman-Liebermann, B.; Liebermann, D.A.  
Nucleic Acids Res. 18, 2823, 1990  
A:Title: Sequence of MyD116 cDNA: a novel myeloid differentiation primary response gene  
A:Reference number: S10001; MUID:90251472; PMID:2339071  
A:Accession: S10001  
A:Molecule type: mRNA  
A:Residues: 1-657 <OR>  
A:Cross-references: UNIPROT:P17564; EMBL:X51829; NID:G53040; PIDN:CAA36128.1; PID:G53041

```
Query Match          5.5%; Score 204; DB 2; Length 657;
Best Local Similarity 21.5%; Pred. No. 4.9e-06;
Matches 149; Conservative 86; Mismatches 290; Indels 168; Gaps 27;

Qy 60 LLSQALLPSLFQKLLWSQSLGIPTRMLDFAASYALRASGRRESDAPTQKSL 119
Db 21 LLSPLMGLLSR-----AWSRLRGPEVPEAWL-----AKTVTGADQ 55

Qy 120 YTAAGLFAKTRVSTLALARG-----GTPVAVLVLRLEVKLKAQERAL---DSPAFTLL 171
Db 56 IEAALLNPTPVSGNL-LPHGETESGSP-----EQSQAORLCVLEAESPP----- 102

Qy 172 EQQLMGVELLPS-SIQAGLVSHRELD-SSSGPLSVQSLG-----NFKVSVY---LLN 219
Db 103 --ETWGLSNVDENYNAKPQDDLREKEMERTAGKATLQAGLQGDKRLGEVVAEEGVAE 160

Qy 220 PSYLDYLPQGLRCOSSAGGQGVCFRTLTPESCVLSYSDGCHPOPLRAEMSATAWRCPP 279
Db 161 PAY-----PTSQLEGGPAEN-----BEDGETVTKTYQASAASIAFGVKPS 199

Qy 280 LSTEGLPEIHRRRWLVFLPNQODLPTLDQNGVHSLSEHNLLRMDDPHOCTDNPQA 339
Db 200 TPVPFLGAEHQATE-----EKGTEKADPSNPSGSHSRAWEY--YSREKPKQSGEAKVE 254

Qy 340 AVSPAADRP-----EPTKKPLVIEQVSQSPGSSLFCELPVEKECEDHTNATDLSDRG 395
Db 255 AHRAGQGHPCRNAEAEAGGPETTF--VCTGNALFKAWYVRPGEDTEEDNDSDSASEDT 312

Qy 396 ESLPVSTRPVCSNKLIIDYILGAPSDLEASSDSSED-----W-----GE 435
Db 313 AQTGATPHTSAFLKAWYVRPGEDTEEDSDSASEDTAQTGATPHTSAFLKAWYVRPGE 372

Qy 436 EPEDDGFSDGSLSESDVEQSEGHLWNSFHSVDPYPKQNFATIQTAARIAPRDPDS 495
Db 373 DTEENSOLDL--AEEDTAQTGATHT--SAFLKAWYVRPGEDTEEDNDSASEDTAQT 429

Qy 496 GTSSGSGCGVSCQEGPLPETPDHSSGEEDDWEPSADEAENLKLWNSFCHSEDPYNLNF 555
Db 430 GATPHTSPFLKAWYVRPGEDTEDTEEDSENVAAGDSEADSSQSPC----- 478

Qy 556 KAPFQPS-----GKNWKGQDSKASSEVTVAFSGHH-----TLSCKAQLLS-- 598
Db 479 ---LQPRCLPGEKTKGRGEPPLFQVAFYLFGEKPESPFAAPKLPLRLQRLRLFKAPT 535

Qy 599 --QEDNCPGCGLGEALAGERYTHKRVKVTLEEVEYYIS-----GDEDKRGWEEFAR 651
Db 536 RDQPEIP-----LKARKVHFAEKVTVHFLVWAGPAQAARRGPEWGFAR 580

Qy 652 DGCGRFQKRIQETEAIGVCLAFEHREKMFNRLR 684
Db 581 DRSRFARIIAQAEKLGPLYLPDSEARAWARLR 613
```

## RESULT 2

A56535

gadd34 protein - long-tailed hamster  
N:Alternate names: MyD116 protein homolog  
C:Species: Cricetus longicaudatus (long-tailed hamster)  
C:Date: 19-Oct-1995 #sequence\_revision 19-Oct-1995 #text\_change 09-Jul-2004  
C:Accession: A56535  
R:Zhan, Q.; Lord, K.A.; Alano Jr., I.; Hollander, M.C.; Carrier, F.; Ron, D.; Kohn, K.W.  
Mol. Cell. Biol. 14, 2361-2371, 1994  
A:Title: The gadd and MyD genes define a novel set of mammalian genes encoding acidic pro-  
tein kinases  
A:Reference number: A56535; MUID:94187707; PMID:8139541  
A:Accession: A56535  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-590 <ZHA>  
A:Cross-references: UNIPROT:Q60465; GB:L28147; NID:G452489; PIDN:AAA36983.1; PID:G452490  
C:Genetics:  
A:Gene: Gadd34

```
Query Match          4.6%; Score 171.5; DB 2; Length 590;
Best Local Similarity 18.1%; Pred. No. 0.00069;
Matches 100; Conservative 68; Mismatches 198; Indels 187; Gaps 18;

Qy 250 PESCYLSDGCHPOPLRAEMSATAWRCPP 292
Db 103 PETLGLSDDDKQGDGPREQGRAHTAGLPILLSPGLQADKSLGEVVAGEGVTETAYPT 162

Qy 293 MRVLVFLPNQGDPLPTLDQNG-----YHSLSEEH 323
Db 163 SHW-----EGCPSEEDGETVKAFRASADSPGHKSSTSVYCPGAEHQATEKQ 213

Qy 324 NLLRMDP-----QHCTDNPAAVSPAADREP-----TEKKPELV 358
Db 214 TENKADPPSPSGSHSRAWEYCSKQGEA-----DPEPHRAGKYQLCQNAEAESEAK 267

Qy 359 IOEVQSPOQSSSL--FCELPVEKECEDHTNATDLSDRGESLPVSTPVCNKLIDYILG 416
Db 268 VSSLVSSVSGNAFLKAWYVRPGEDTEDDDSDWGSABEEGKALSSPTSPE-----HDFLKA 322

Qy 417 GAPSDLEASSDSSESDGEPDDGDFSDGSLSESDVEQSEGHLWNSFHSVDPYPKQPN 476
Db 323 WYVRPGEDTEDDDSDWGSABEE-----EGKALSSPTSPEHDFLKA-----VYRPG 369

Qy 477 FTATIQTAARIAPRDPDSGTSSGSGVSCQEGPLPET----- 516
Db 370 DT-----EDQDSDWGSABEKGQLAQTATPHTSAFLKTVWCCPGEDTEDDDC 416

Qy 517 ----PDHSSGEEDDWEPS-----ADEAENLKLWNSFCHSEDPYNLLNFKAPFPQSGK 564
Db 417 EVVVPEDSEAADPKSPSHEAQGLPGEQTEGL-----VEAHSFLQVAFYLPGEKPA 470

Qy 565 NWKGRQDSKASSEVTVAFSGHHILLSCKAQLLSQEDNCPGCGLGEALAGERYTHIKRK 624
Db 471 PW-----TAPKLPLRLQRLTLLRPTQ---DQDPETP-----LRARK 505

Qy 625 VTFLEEVEYYIS-----GDEDKRGWEEFARDCRFQKRIQETEAIGVCLAFEHREK 679
Db 506 VHFSENVTVHFLVWAGPAQAARRGPEWQLARDRSRPARIIAQAEKLGPLYLTPAFRARA 565

Qy 680 FNRLRLESKDLLL 692
Db 566 WARLGNPSFLAL 578
```

## RESULT 3

T33100  
hypothetical protein H11E01.3 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 09-Jul-2004  
C:Accession: T33100  
R:Geisel, C.; Wansley, P.  
submitted to the EMBL Data Library, May 1998  
A:Description: The sequence of C. elegans cosmid H11E01.  
A:Reference number: Z21282  
A:Accession: T33100



R;Ishikawa, K.; Nagase, T.; Nakajima, D.; Seki, N.; Ohira, M.; Miyajima, N.; Tanaka, A.; submitted to the EMBL Data Library, October 1997  
A;Description: Prediction of the coding sequences of unidentified human genes. VIII. The A;Reference number: Z14082  
A;Accession: T00062  
A;Status: preliminary; translated from GB/EMBL/DBDJ  
A;Molecule type: mRNA  
A;Residues: 1-1571 <ISH>  
A;Cross-references: UNIPROT:O43161; EMBL:AB007894; NID:G2662148; PIDN:BAA23707.1; PID:G262148  
A;Experimental source: brain; clone HH2165  
C;Genetics:  
A;Note: KIAA0434

```
Query Match          3.7%; Score 136.5; DB 2; Length 1571;
Best Local Similarity 23.9%; Pred. No. 0.66;
Matches 140; Conservative 68; Mismatches 198; Indels 179; Gaps 36;

Qy 61 LSQLALLPSLFQKLLWSQSGGLIPTRWLD--FAASY-----SALRA 102
Db 424 LSQLVSRQPPKSPQ-VLYSPVS-PLSPHRLDTSFASERLNKAHVSPQKHTADSALR- 480

Qy 103 SRGRESAP--TVQKSL-----YTAAGLFAKTRVSTLALARGTTPVAVLRLVLEVL 155
Db 481 ---QOTLPRPMKTLQRLSLDKPLSPPTAESAKERF--SLYQHOGGLGQVSALPPNSLV 535

Qy 156 KAQERALDSAP-----TFLLEQOLAGVLLPSSLOAGLVSHRELDSS 198
Db 536 RKVKRTLSPPEEAEHLPLAQASPOLYASLLQGLTPTVTPAT-KASLL--RELDR- 591

Qy 199 SSGPLSVQSLGNFKVSVYLLN-----PSYLDYLPQLGL--RCOSSA--- 237
Db 592 -----DLRLVEHESKLRKKQAEDEEKEIDAKLYL-ELGITQKESLAKDR 639

Qy 238 GGGQFVGFRITLTPSCYSE-----DGC-HQOPLRAEMSAWARRCPPLSTEGUPEI 288
Db 640 GGRDYPPLRLGLGHRDYLSDELNLQRLQGCTTPAGQVDFPATA--AAPATPSGPTAF 696

Qy 289 HHRMRWLVLQP-----NQODLPTLDQDQNGHSLSEEHLLRMDPOHCTDN----- 336
Db 697 QQPR-----FQPPAPQVSAGSGPT-----PAHQAPTVPGSTY 734

Qy 337 PAQAVSPAADRPETKEKPELVIQ-----VSQSPQ-----GSSLFCFELVEKECEDH 385
Db 735 PAPAPPPGASYP-----AEPGLPNQAQFRTPGHYAGQTPMTTQSTLFPVPADSRAPLQK 789

Qy 386 TNATDLSDRGSLPVSTRPVCNKLIDYLGGAPSDLEASDSSESDW--GEBPEDDGF 443
Db 790 PROTSIADLEQKVPNTYEVIAAPV---PMSSAPSETSYSGPAPVSSGYEQGKVPVPRAG 846

Qy 444 SDGSLSESDVEQDSGLHLWNSFHSVDYKPKQNF-----TATIQTAARIAPR 490
Db 847 DRGSVSGPAPTYPSDSH-----YTSLEQNVPRNYVMIDISBELTKDSTAPDSQRLPEL 902

Qy 491 DPSDSGTSWS--GSCGVGSCQEGP-LPPTPDHSSGEEDDWEPSSAD 532
Db 903 GPGSSGRPGKEPGEV---LDGPTLPC--YARGESESESDY 942
```

RESULT 6  
C88042  
protein F56D12.4 [imported] - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 10-May-2001  
C;Accession: C88042  
R;anonymus, The C. elegans Sequencing Consortium.  
Science 282, 2012-2018, 1998  
A;Title: Genome sequence of the nematode C. elegans: a platform for investigating biology  
A;Reference number: A75000; PMID:9851916  
A;Note: see websites genome.wustl.edu/gsc/C\_elegans/ and www.sanger.ac.uk/projects/C\_elegans/ and  
A;Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and  
A;Accession: C88042  
A;Status: preliminary  
A;Molecule type: DNA

A;Residues: 1-527 <STO>  
A;Cross-references: GB:chr\_II; PIDN:AB66118.1; PID:G2315675; GSPDB:GN00020; CESP:F56D12.4  
C;Genetics:  
A;Gene: F56D12.4  
A;Map position: 2

```
Query Match          3.6%; Score 135; DB 2; Length 527;
Best Local Similarity 21.3%; Pred. No. 0.18;
Matches 112; Conservative 57; Mismatches 178; Indels 180; Gaps 27;

Qy 96 SYSY-----LRASRGRESDAPTQKSLSYTAAGLFAKTRVSTLALARGTTPVAVLRLV 150
Db 33 SYSADWLHNDNSGGDDSEEDSDAEFLVHPCLRSIIKSK--SAYAVPRNSDDVAAMAAM 90

Qy 151 LEVKKLQAE-----RALDSAAPTFLRQOLMGVLLPSSLOAGLVSHRELDSSSGP 202
Db 91 KNNRIKPSKTSFETSLSKMMESTSKSABEAESVADLLPT-----LDEKWPDP 139

Qy 203 LSVQSLGNFK-----VVSYLNPYSYLDYLPQGLRCOSSAGGQFVGFRITLTPESCYL 255
Db 140 --EENENLKSRRKSNIMRY--SPSLCDSLDLSNLHDKDSEDEDLINR--PSGCWS 192

Qy 256 SEDGCHPOPLR--AEMSAWARRCPPLSTEGLPPIHHRMRWLVLQPNQODLPTL--- 310
Db 193 SPNPFDETRGVSPMNQKCM-----SAPDLEDIVFKTNCVLYNYDVIVGGKIPILREL 246

Qy 311 -DQDN---GYHSLEEEHLLRMDPOHC-----TDNPAQAVSPAADRPETKEK 354
Db 247 IDSDIAIGYFVETPDN-----DPSVCSYRLPSGLSAPSPSPRPLSPVFGQAPPOYR 302

Qy 355 PELVIQVSOS---PQSSSLFCFELVEKECEEDHTNATDLSDRGESLPVSTRPVCNKL 411
Db 303 -----MVSVSGMLPRPPDIEAF---ADQNDYQ----- 330

Qy 412 DYILGGAPSDLEASDSSESDWGEEDDGFDSGSLSESDVEQDSGLHLWNSFHSVDP 471
Db 331 -----APPTFNIGSDLEDNITEEBEDBED-----EEDVRKTEE---VW----- 367

Qy 472 YKQNFAT-----IQTAA-----RIAPRDPDSGTSWSGSCGVGSCQEGPLP--- 514
Db 368 --PENATANFMSDGVRRVHRSATFQINRYLDDDDVTGGT-----VSPPLPRIT 414

Qy 515 -----ETPDHSSGEEDDWEPSSADEAENLKLWNSFCHS 546
Db 415 VSRSAHSQLYNLGAETEDDADGGDD---DDDVAENL---NENCSS 455
```

RESULT 7  
T32155  
hypothetical protein C10F3.1 - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 09-Jul-2004  
C;Accession: T32155  
R;Davidson, S.; Wohldmann, P.  
submitted to the EMBL Data Library, September 1997  
A;Description: The sequence of C. elegans cosmid C10F3.  
A;Reference number: Z21127  
A;Accession: T32155  
A;Status: preliminary; translated from GB/EMBL/DBDJ  
A;Molecule type: DNA  
A;Residues: 1-782 <DAV>  
A;Cross-references: UNIPROT:O16883; EMBL:AF022968; PIDN:AAB69883.1; GSPDB:GN00023; CESP:  
A;Experimental source: strain Bristol N2; clone C10F3  
C;Genetics:  
A;Gene: CESP:C10F3.1  
A;Map position: 5  
A;Introns: 52/3; 107/2; 224/1; 318/2; 378/2; 455/1; 657/2

```
Query Match          3.6%; Score 133.5; DB 2; Length 782;
Best Local Similarity 20.9%; Pred. No. 0.4;
Matches 108; Conservative 67; Mismatches 188; Indels 153; Gaps 23;

Qy 222 YLDYLPQGLRCOSSAGGQ--FVGFRITLTPESCYLESDGCHGPQPLRAEMSAWARRCPPL 280
```



Db 285 WADLANIGORPPSPDQAQGFENFROFTNDLCRIGD--CMLDCIRSKFNT----RCEGS 338  
Qy 281 STEGLPIHHR---RMRWLVLQPNQODLPTLDQDNGYHSLSEHNLLRMD----- 329  
Db 339 AGTLLSEVFRPIAATONKLSILRPLGTMP---EOCGYLTNNAELKKHRIATWDEEL 395  
Qy 330 -----POHCTDNPAQ-----AVSPAADRPEPEKK-PELVIOE----- 361  
Db 396 KRWYAEKIAKEARDTAQDEILANLVLPLDENGVLPRALPELKSIESPLDVSKVLQLLI 455  
Qy 362 -----VSQSPQSSSLFCELPVKECEED-----HTNATDLSDRGESL 398  
Db 456 LDMYSNNKTBELNISEKNVTSTFSE-PSEKEDASTVTSVPLHTNATD----- 506  
Qy 399 PVSTRPVCSNKLIDYI-----LGGAPSDLEASSD--SESEDWGEBEPEDDG- 441  
Db 507 -----SEILEHISEKSTEESGSGSGEMSGDGSNEASGEGSGYDASGSGDNGSE 557  
Qy 442 FDSGSLSESDVEODSEGLHLNWSFHSVDPYKPNQNTATTQTAARIAPRDPDSGTSWSG 501  
Db 558 FNSGSGSEASEGEGSGSDQSGN---YK---MIESIESGEPFSGSGEGSGDTASS 610  
Qy 502 S-----CGVSCQEGPLPETHSSGER-DDWEPADAEANLKLWNSFCHSEDPY 550  
Db 611 DTSIDDKSIIRSGGSAES--VSEILOEASGEDAPTLP7SESTGYKIDHS----- 660  
Qy 551 NLLNFKAPFPGSKNWKGRQDSKASSVTVAFSGHHTLLSCAKLLSQSDNCPG----- 605  
Db 661 ---GFGESSGSGSIELRDSGEGSAEYDASGSGD-----NSGDFNSSGSGGEA 707  
Qy 606 CGLGEALAGERYTHIRKKVTFLEEVTEYISGDED 641  
Db 708 SGVSGSGSDQSGNKKIEVIESGDIYFSGSSN 743  
RESULT 8  
brevican precursor - rat  
N:Alternate names: aggrecan-like protein  
C:Species: Rattus norvegicus (Norway rat)  
C>Date: 01-Feb-1995 #sequence\_revision 12-May-1995 #text\_change 09-Jul-2004  
R:Accession: S49126; 155457  
R:Seidenbecher, C.I.; Langnese, K.; Wax, H.; Seidel, B.; Garner, C.C.; Gundelfinger, E.D. submitted to the EMBL Data Library, June 1994  
A:Description: Molecular cloning of a new member of the aggrecan/versican family of proteoglycans  
A:Reference number: S49126  
A:Accession: S49126  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-883 <SEI>  
A:Cross-references: UNIPROT:P55068; EMBL:X79881; NID:G509396; PIDN:CAA56255.1; PID:G509397  
J. Biol. Chem. 270, 27206-27212, 1995  
A:Title: Brevican, a Chondroitin Sulfate Proteoglycan of Rat Brain, Occurs as Secreted and Membrane-associated Forms  
A:Reference number: 155457; MUID:96070828; PMID:7592978  
A:Accession: 155457  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-883 <RES>  
A:Cross-references: EMBL:X79881; NID:G509396; PIDN:CAA56255.1; PID:G509397  
C:Comment: For an alternative splice form, see PIR:A53908.  
C:Superfamily: aggrecan; C-type lectin homology; complement factor H repeat homology; EGF repeats  
C:Keywords: alternative splicing  
F:49-138/Domain: immunoglobulin homology <IMW>  
F:173-250/Domain: link protein repeat homology <LNK1>  
F:271-352/Domain: link protein repeat homology <LNK2>  
F:626-657/Domain: EGF homology <EGF>  
F:664-784/Domain: C-type lectin homology <LCH>  
F:791-847/Domain: complement factor H repeat homology <FHD>

Query Match 3.6%; Score 133.5; DB 2; Length 883;  
Best Local Similarity 22.2%; Pred. No. 0.47;

Matches 154; Conservative 63; Mismatches 217; Indels 259; Gaps 40;  
Qy 140 GGTFAVLV---LRLEVKLKAQERALDSAAFTLLEQQQLMGVELLP--SSLQAGLVSHRE 194  
Db 83 GDREVEVLVARGLRVKVNEAYRFRVALPAYPASLTDVSLVLSLRPNDSGVYRCEVQHG 142  
Qy 195 LSSSSGPIUSVQSL-----GNFKVSVYLLNFSYLDYLPQLGLRCQ 234  
Db 143 DSSDAVEYKVGWVFLYREGSARYAFSAGAEACARIGARIATPEQL-YAAYLGGYEQ 201  
Qy 235 SSAG--GGQFVGPRTLTP--ESCVLSGDG-----CHPOPLRAEMSA- 271  
Db 202 CDAGWLSQDQVRYPIONPREACYGDMGDPGVGVNRYGVGDDLYDYVYCAEDLNGELFLG 261  
Qy 272 -----TAMR-----RCP-----PLSTE----- 283  
Db 262 APPGKLWTBEARDYCLERGAQIASTQQLYAAWNGGLDRCSPGWLAGDSVRYPIITPSQRC 321  
Qy 284 --GLPEIHRRMRWLVLQPNQODLPTLDQDNGYHSLSEHNLLRMDPOHCTDNPA--- 338  
Db 322 GGLPGVK-----TLFLFPNQ-----TGFPKQNRFNV-----YCFRDSAHPS 359  
Qy 339 ---QAVSPAADRPEP---TEKKPELVI-QEVSQSPQSGSLFCELPVKECEDEHTNATD 390  
Db 360 AFSEASPSADGLEAIVTTEKLEELQLQEAVERSESRAIY-SIFI-----TE 407  
Qy 391 LSDRGESLPVSTRPVCSNKLIDYLGAP--SDLEASSDS-----ESEDWGEPEDDGFD 443  
Db 408 DGGGSGSTPEDP-----AEAPRTPLESETQSVAPPTGSESEGEALEBERF 454  
Qy 444 SDGSLSESDVEQSGEGHLW-----NSFHSVDYKPNQNTATTQTAARIAPRD 491  
Db 455 KUTETPKKEEQ--ENLWVMTLSPLTGLETHSLSQVSP--AAVLQVQASPSRP 511  
Qy 492 PSDSGTWSGSCGVGSCQ---EGPLPETPHS---SGEEDDWEPSEADEANLKLWNSFCH 545  
Db 512 PRVHGPT-----VETLQPPGSGLSTPTDGAREVGGTGPSELGVPREREAGSS--S 563  
Qy 546 SEDPYNLLNFKAPFPQSG-----KNMKGR-----QDSKASSEVTVA 582  
Db 564 LEDGPSLL--PETWAPVGTREVETPSEKSGRTVLGTSGVQAQVPLPTDSASRGVAVAP 621  
Qy 583 SGHTLLS-CK--AQLESQED-NC---PG-----CGLG-----EALAGERYTHIR 622  
Db 622 SSGDCIPSPCHNGTCLKEEGEPRCLCVPGYGGDLCDVGLHFCSPGWEPFGQACYKHFST 681  
Qy 623 KKVTFLEEVTEYISGDEDRKGPWEERFARDGCR 655  
Db 682 RR-----SWEE-AESQCR 693

## RESULT 9

P90603  
vipe-like (mycoplasma hyorhinis) lipoprotein [imported] - Mycoplasma pulmonis (strain UAB C)  
C:Species: Mycoplasma pulmonis  
C>Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 09-Jul-2004  
C:Accession: P90603  
R:Chambaud, I.; Heilig, R.; Ferris, S.; Barbe, V.; Samson, D.; Gallieson, F.; Moszer, I.; Nucleic Acids Res. 29, 2145-2153, 2001  
A:Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma pulmonis  
A:Reference number: A99512; MUID:21267165; PMID:11353084  
A:Accession: P90603  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-682 <KUR>  
A:Cross-references: UNIPROT:Q98PT8; GB:AL445566; PID:gl14090149; PIDN:CAC13907.1; GSPDB:G13907.1  
A:Experimental source: strain UAB CTIP  
C:Genetics:  
A:Gene: MYPV 7340  
A:Genetic code: SGC3

Query Match 3.6%; Score 132.5; DB 2; Length 682;  
Best Local Similarity 20.6%; Pred. No. 0.39;





Qy 502 SCGVSCQEGPLPETPDH-----SSGEEDWEPsADEARNLKLWNSFCHSDPYNLN 554  
Db 254 SSSDSSSDSTSSDSDSDSDSGSSSELTKEATADES----- 293  
Qy 555 FKAPPQPGKWKGRQDSKASEVTVAFGSHHTLLSCAKQLLESQEDNCPGCGLGEALAG 614  
Db 294 -KAETPASNESSTPSASSSS-----ANKLNIP-AGTDEIKEG 330  
Qy 615 ER--YTHIKRKKVTF-LREVTEYISGDEDRKGPWEFARD 652  
Db 331 QRKHSRVDRSKINFEAWELDTNTYK---AAGTWGERANE 368  
RESULT 14  
T03852  
protein phosphatase 1 binding protein spinophilin - rat  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 09-Jul-2004  
C:Accession: T03852  
R:Allen, P.B.; Oulmet, C.C.; Greengard, P.  
Proc. Natl. Acad. Sci. U.S.A. 94, 9956-9961, 1997  
A:Title: Spinophilin, a novel protein phosphatase 1 binding protein localized to dendric  
A:Reference number: Z15121; MUID:97420791; PMID:9275233  
A:Accession: T03852  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-817 <ALL>  
A:CROSS-references: UNIPROT:O35274; EMBL:AF016252; NID:g2462850; PIDN:AAB72005.1; PID:g2462850

Query Match 3.4%; Score 126; DB 2; Length 817;  
Best Local Similarity 22.8%; Pred. No. 1.4;  
Matches 89; Conservative 44; Mismatches 134; Indels 124; Gaps 19;  
Qy 220 PSYLDYLPQGLRCOSSAGG--QVGFRTLTPECYLSE-----D 258  
Db 141 FSLQETKRLPERSVPAASGGDKAVARRLLRQERASLDKRLDVVVRFNFGSTEALDKLD 200  
Qy 259 GCHPOPLRAEMSA-----TAWRRCF-PLSTGLPEIHRRM--RWLVFLQPNQGD 307  
Db 201 ADASVPTVSQLSAYPEKADSTGLHRAFGPPAAGAPQVNSKLVTKRSRVQPP----- 254  
Qy 308 PTLDDNGYHLSLEEHNLRLMDPOHCTDNPAQAVSPAADRPEPTE--KKPELVTOE----- 361  
Db 255 PPPAPSGDAATEKDRGPGGQP-----PQHRVAPARPPKPREVRKIKPVEVEESGES 308  
Qy 362 VSQSPQSSSLFCELPVEKECEDHTNATDLS-----DRGESLP-----VST--RPVCS 407  
Db 309 EASAPGVEIQAEVTVHAALENGSTATTATAPAEPEPKAEAVPEEASSSVATLERGVDN 368  
Qy 408 NKLDYILGGAPSDLESSESDWGEPEDD-----GFDSDGSLSESVDVEQDSGL 460  
Db 369 GRAPDM---APEEVD---ESKEDFSEADLVDSAYSGLGEDSGGSALEEDDEEB-- 419  
Qy 461 HLWNSFHSDVPYKPNFTATTQTAARIAPRDPDSGTSGVSCGVSCQEGPLPETPDHS 520  
Db 420 -----DGEPPYEPESGVEIPGLSE-----EEDPAPSRKIHF 451  
Qy 521 S-----GEED-----DWEPsADEAE 535  
Db 452 STAPIQVFSTYSNEDYDRRNEVDVDPMAASAE 482

RESULT 15  
T46116  
hypothetical protein T2J13.20 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 04-Feb-2000 #sequence\_revision 04-Feb-2000 #text\_change 09-Jul-2004  
C:Accession: T46116  
R:Rieger, M.; Gabel, C.; Mueller-Auer, S.; Schaefer, M.; Zipp, M.; Mewes, H.W.; Mayer, K.  
submitted to the Protein Sequence Database, November 1999  
A:Reference number: Z23023  
A:Accession: T46116  
A:Status: preliminary

A:Molecule type: DNA  
A:Residues: 1-1229 <RIE>  
A:CROSS-references: UNIPROT:O9SMV0; EMBL:AL132967  
A:Experimental source: cultivar Columbia; BAC clone T2J13  
C:Genetics:  
A:Map position: 3  
A:Introns: 683/1; 723/1; 769/2; 842/3; 877/1; 899/3; 908/1; 961/3; 1016/3; 1118/3; 1169/3; 1169/3; 1169/3  
A:Note: T2J13.20

Query Match 3.4%; Score 126; DB 2; Length 1229;  
Best Local Similarity 22.4%; Pred. No. 2.4;  
Matches 120; Conservative 63; Mismatches 209; Indels 144; Gaps 29;  
Qy 184 SLOAGLVSHRELDSSSSG-----PLSVQSLGNFKVVSYLINPSYLDYLPQLG 230  
Db 730 SSHAKFVEYRTAPSSSEGGNCFHYASRRVFPQRIHHIDGSGFLKY-----NSDYITRKH 784  
Qy 231 LRCOSSAGGQGFVGRFRTLTPESCYLSDEGCHP-OPLRAEMSATAM--RRCPPLSTEGLEPE 287  
Db 785 LRKNTQATAAYYVDSAS-DPEK-QTGKSRVHPSEIRASLPQNDGDSRLSPAETTRTIIIE 842  
Qy 288 IHHRMRWLFLQPNQGDLPDLDQDNGYHLSLEEHNLRLMDPOHCTDNPAQAVSPAADR 347  
Db 843 VN-----NKGTMLTSGSIGDVH-----ENILWPDPIYITDQGNLYFQV--- 882  
Qy 348 PEPEKKPELVIOEVSQSPQSSILFCELPVEKECEDHTNATDLSDRGESLPVSTRPVCS 407  
Db 883 -----KEDEDVMQSVTSNNYVLL-----KHVSDDVMDALQQVIVGFDTEM 925  
Qy 408 NKLDYILGGAPSDLEA-----SSDSESDWGE-----EPEDGFDSDGSLSES 452  
Db 926 IKEME-LMGLSDSDFETEDDESDGDDSDTGEDEDEEWVAILEDDEDDDDDDDDDD 984  
Qy 453 VEODS-EGHLWNSFHSDVPYKPNFTATTQTAARIAPRDPDSGTSGVSCGVSCQEG 511  
Db 985 DSDSDSLGDMANLETMRSCHPMFFA---KRTEVASNDPVD---WMDQPSAGLAIQG 1037  
Qy 512 PLPE--TPDHSS-----GEEDWEPsADEAENL--KLWNSFCHSESDPYNLNFKAPFQPS 562  
Db 1038 LLSHILVEDISDIQKLADSNSTINGNKDAENLVKL-----ED-----NSKAGGDES 1085  
Qy 563 GKWKGRQDSKASEVTVAFGSHHTLLSCAKQLLESQEDNC-----PCGGLGEA 611  
Db 1086 --EIDSSQDEKARN--VVAFA---YKLEIRIQLITAQGDQTEVEVEDVRKAQP-----DA 1133  
Qy 612 LAGERYTHIKRKKVTFLEEVTEYISGDEDRKGPWEFARDGCRFKRQIOETVAI 667  
Db 1134 IA-----HASAEIISRLSE-----SGDK-----ITALKSLCWRHNSIQAEVKL 1173

Search completed: September 16, 2005, 10:46:39  
Job time : 20.8086 secs



```

RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
RA Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
RA Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,
RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.,
RA Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
RN
[7]
RN SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Spinal cord;
RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
RA Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,
RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
RA Katoh H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,
RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,
RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
RA Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
RA Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,
RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.,
RA Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
RL EMBL; AK049028; BAC33517.1; -
DR EMBL; AK082957; BAC38708.1; -
DR MGI; MGI:2444211; Ppplr15b.
DR GO; GO:000164; C:protein phosphatase type 1 complex; IDA.
DR GO; GO:0004722; F:protein serine/threonine phosphatase activity; IDA.
DR GO; GO:0045182; F:translation regulator activity; IC.
DR GO; GO:0006983; P:ER-overload response; IDA.
DR GO; GO:0006446; P:regulation of translational initiation; IC.
DR GO; GO:0042542; P:response to hydrogen peroxide; IMP.
DR GO; GO:0006979; P:response to oxidative stress; IDA.
DR Hypothetical protein.
SQ SEQUENCE 697 AA; 77711 MW; E439B12615F3737 CRC64;
QY Query Match 93.3%; Score 3460.5; DB 2; Length 697;
DB Best Local Similarity 94.3%; Pred. No. 8e-204;
M 658; Conservative 7; Mismatches 32; Indels 1; Gaps 1;
QY 1 METGTHRAKRPGLGFWFLPFLRRSHACSFPPPPSSRQNGNSALPERRTRYWTKL 60
DB 1 METGTHRAKRPGLGFWFLPFLRRSHACSFPPPPSSRQNGNSALPERRTRYWTKL 60
QY 61 LSQALLLPSLFQKLLWSQLSGGLIPTRWLDFAASYSALRASGRDESAPTVQKLSY 120
DB 61 LSQALLLPSLFQKLLWSQLSGGLIPTRWLDFAASYSALRASGRDESAPTVQKLSY 120
QY 121 TAAGLFAKTRVSTLALARGSTPVALVLRLEVKLKAQERALDSAAFTLEQLQWVGL 180
DB 121 LRLD-SSDLVVSLSLWLEGLQWQCSSDLELKLKAQERALDSAAFTLEQLQWVGL 179
QY 181 LPSSLQAGLVSHRELDSSSGPLSVQSLGNFKVSYLLNPSYLDYLPQLGLRCQSSAGG 240
DB 180 LPSSLQAGLVSHRELDSSSGPLSVQSLGNFKVSYLLNPSYLDYLPQLGLRCQSSAGG 239
QY 241 QFVGFRLLTPESCVLSDGCHPOPLRAEMGATAWRRCPPLSTEGLPPIHRRRWLVFLQ 300
DB 240 QFVGFRLLTPESCVLSDGCHPOPLRAEMGATAWRRCPPLSTEGLPPIHRRRWLVFLQ 299
QY 301 PNCQDLPDLDDQNGVHSLSEENLLRMDQHCNTDPAQAVSPAADRPETPEKPELVQ 360
DB 300 ANKQQLPTDQNGVHSLSEENLLRMDQHCNTDPAQAVSPAADRPETPEKPELVQ 359
QY 361 EVSQSQGSSLFCELPVKECEEDHTNATDLSRGSPLVSTRPVCNKLDYILGAPS 420
DB 360 EVSQSQGSSLFCELPVKECEEDHTNATDLSRGSPLVSTRPVCNKLDYILGAPS 419
QY 421 DLEASSDSESDGEEPEEDGDFSDGSLSDVVDQSEGLHWNSTFSDVPYKQNTAT 480
DB 420 DLEASSDSESDGEEPEEDGDFSDGSLSDVVDQSEGLHWNSTFSDVPYKQNTAT 479
QY 481 IQTAARLAPDPSDGTSGVSGCGVCGQGPETPDHSSGEDDWEPSADEENLKW 540
DB 480 IQTAARLAPDPSDGTSGVSGCGVCGQGPETPDHSSGEDDWEPSADEENLKW 539

```



```

Qy 541 NSPCHSDPYNLLNFKAPFPGSGKNWGRQDSKASSEVTVAFSGHHTLLSKAQLLESQE 600
Db 540 NSPCHSDPYNLLNFKAPFPGSGKNWGRQDSKASSEATVAFSGHHTLLSKAQLLESQE 599
Qy 601 DNCPCGGLGALAGERYTHIRKRVKVTLEEVTEYIISGDDRKGPWEFEFARDGCRFOKRI 660
Db 600 DNCPCGGLGALAGERYTHIRKRVKVTLEEVTEYIISGDDRKGPWEFEFARDGCRFOKRI 659
Qy 661 QETEVAGYCLAFEHREKMFNRLRIESKDLLLYSNVKK 698
Db 660 QETEVAGYCLAFEHREKMFNRLRIESKDLLLYSNVKK 697

RESULT 2
Q6P156 PRELIMINARY; PRT; 713 AA.
AC Q6P156;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DE Protein phosphatase 1, regulatory subunit 15B.
GN Name=PPP1R15B;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP TISSUE=Eye;
RC MEDLINE=2477932; PubMed=12477932; DOI=10.1073/pnas.242603999;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.H., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haieh P.,
RA Diatchenko L., Narusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Brownstein M.J., Uedin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A.C., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywicki M.I., Stalska U., Smalls D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Eye;
RA Director MGC Project;
RL Submitted (JAN-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC065280; AAB65280.1;
SQ SEQUENCE 713 AA; 79125 MW; 76C2DA38F976A85 CRC64;

Query Match 59.7%; Score 2214; DB 2; Length 713;
Best Local Similarity 64.8%; Pred. No. 2.2e-127;
Matches 454; Conservative 55; Mismatches 174; Indels 18; Gaps 7;

Qy 1 METGTHARKPFGRLGFWRLPFL-RRSHACSEFPFPPSSQRPN- ----SALPERRT 54
Db 1 NEPTGGSKKLGPAGRFPPPPRRSQAGSKFPFPLGPNESGNPTLLSSAQPETRV 60
Qy 55 RYWTKLLSQLALPFLPKQLLLSQSGGIPTRWLDFAASYALSALRASRGRESDAPTV 114
Db 61 SYWTKLLSQLALPGLLQKLVLSQLPGGMPFTRWLDFAGVYALSALRALKGREKPAAPTA 120
Qy 115 QKSLSYTAAGLFAKTRVVSTLALRGTPVAVLVRLEVKLKAERALDLSAAPTFLLEQQ 174
Db 121 QKLSLSQLD-SSDPSTVSPDLWLEEGTHWQYSPDPLKLELKALGSLDPAQAFLLEQQ 179

```

```

Qy 175 LMGVELLPSSLAQGLVSHRELSSSGPLSGVSLGNFKVSVYLLNPSYLDLPLQLGLRCQ 234
Db 180 LMGVELLPSSLAQGLVSHRELSSSGPLSGVSLGNFKVSVYLLNPSYLDLPLQLGLRCQ 239
Qy 235 SSAGGQGVFPRLTTPESCYLSDGCHPPLRAEMSATARRCPPLSTEGCLPIHHRMR 294
Db 240 NSDGNSEVVFQTLTPESCLREDHCHPPLSABLIPASWGCPPLSTEGCLPIHHRMR 299
Qy 295 WLVEL-QPNQGDPLTDQNGYHSLSEEHNLRLMDPQHCTDNPAAQAVPAADRP- ---E 349
Db 300 RLEFLQAAKGGDLPTPDQNGYHSLSEEHNLRLMDPKHCRDNPQVPAAGDIPGNTQE 359
Qy 350 PTEKKPELVIEV- ----SOSPGSSLPFCELPVEKECEEDHTNATLSDRGSGLPVSTRP 404
Db 360 STEEKIELTTEVPLALEBESSEGCPSSEIPMEKEPGEGRISVVDVSYLLEGDLPIARP 419
Qy 405 VCSNKLIDYILGAPSDLEASSDSESDMGEERDDGDFSDGSLSDSDVDEQDSEGLHLWN 464
Db 420 ACSNKLIDYILGASSDLESTSSDFEGEDWDEAEDDGFSDSLSDDLQDPEGLHLWN 479
Qy 465 SFHSVDYPKQNPFTATIQTAARIAPRDPDSGTWSGCGV-GSCQEGPLPETPDHSSGE 523
Db 480 SFCSDVPYNPQNFATIQTAARIVPEPSDSEKDLGKSDLENSQSGSLPETPEHSSGE 539
Qy 524 EDDWPSADEAENLKLWNSFCHSEDYDYNLLNFKAPFPGSGKNWGRQDSKASSEVTVAFS 583
Db 540 EDDWESSADEAENLKLWNSFCNSDDPYNPLNFKAPFQTSGENEKGCRDSETPSESIVAS 599
Qy 584 GHHTLLSCKAQLLESQEDNCPGGLGEALAGERYTHIRKRVKVTLEEVTEYIISGDEDRK 643
Db 600 ECHTLLSCKVQLLGSQSECFDSVDJLSSGGRRHVKRKKVTFLEEVTEYIISGDEDRK 659
Qy 644 GPWEFEARDGCRPKQIKQIETEVAGYCLAFEHREKMFNRLR 684
Db 660 GPWEFEARDGCRPKQIKQIETEDAIYCLAFEHREKMFNRLQ 700

RESULT 3
Q96SN1 PRELIMINARY; PRT; 713 AA.
AC Q96SN1;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DE Hypothetical protein FLJ14744.
DB Hypothetical protein FLJ14744.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=14702039; DOI=10.1038/ng1285;
RA Oka T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,
RA Wakamatsu A., Hayaashi K., Sato H., Nagai K., Kimura K., Makita H.,
RA Sekine M., Obayashi M., Nishi T., Shibahara T., Tanaka T., Ishi S.,
RA Yamamoto K., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahari K.,
RA Murakami J., Yasuda T., Iwayanagi T., Wagatsuma M., Shiratori A.,
RA Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E., Omura Y.,
RA Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M., Yamazaki M.,
RA Ninomiya K., Ishibashi T., Yamashita H., Murakawa K., Fujimori K.,
RA Tanai H., Kimata M., Watanabe M., Hiraoa K., Chiba Y., Ishida S.,
RA Ono Y., Takiguchi S., Watanabe S., Yosida M., Hotuta T., Kusano J.,
RA Kanehori K., Takahashi-Fujii A., Hara H., Tanase T., Nomura Y.,
RA Togiya S., Komai F., Hara R., Takeuchi K., Arita M., Inose N.,
RA Musashino K., Yuuki H., Oehima A., Sasaki N., Aotsuka S.,
RA Yoshikawa Y., Matsunawa H., Ichihara T., Shiohata N., Sano S.,
RA Moriya S., Momiyama H., Satoh N., Takami S., Terashima Y., Suzuki O.,
RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,
RA Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,
RA Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,
RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,

```



RA Ono T., Yamada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y.,  
 RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,  
 RA Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,  
 RA Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,  
 RA Togashi T., Oyama M., Hata H., Watanabe M., Komatsu T.,  
 RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,  
 RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuho Y., Yamashita R.,  
 RA Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.,  
 RT "Complete sequencing and characterization of 21,243 full-length human  
 RT cDNAB".  
 RL Nat. Genet. 36:40-45 (2004).  
 DR EMBL; AK027650; BAB5266.1; --  
 DR Genew; HGNC:14951; PP1R15B.  
 SQ SEQUENCE 713 AA; 79125 MW; 26C2D06144AAD25E CRC64;  
 Query Match 59.6%; Score 2213; DB 2; Length 713;  
 Best Local Similarity 64.8%; Pred. No. 2.6e-127; Indels 18; Gaps 7;  
 Matches 454; Conservative 55; Mismatches 174;  
 QY 1 METGTHRAKRPGLGFWFLPFL-RRSHACSSFFPPSSRQNPNGN-----SALPERRT 54  
 Db 1 MEPGTGSRKRLGPRAGRFPPFPFRPSQAGSKFPTPLGPNSENGNTLLSSAQPETRV 60  
 QY 55 RYWKLLSLLALLPSLQKLLMSQLSGGLIPTWLDFAASYSALRASRGRESADPTV 114  
 Db 61 SYWTKLLSLLAPLPLGLLQKVLWSQLFGGMPFTRWLDFAGVYSALRALKGRKPAAPTA 120  
 QY 115 QKSLSYTAAGLFAKTRVVSTIALARGCTPVAVLVLRLEVKLKAQERLDAAPFLLEQQ 174  
 Db 121 QKSLSSQLD--SSDPSTVSPDLDWEEGHHQVSPDPLKLELKAKGSLADPAAQAPFLLEQQ 179  
 QY 175 LWGVELLPSLQAGLVSHRELDSSGSLPSVQSLGNFKVSVYLLNPSYLDLPOLGLRCQ 234  
 Db 180 LWGVELLPSLQSLRSLYNRELSSGSLPGLNQRIDDFSVSVYLLNPSYLDLCPFLVSYQ 239  
 QY 235 SGAGGGQVGRRTLTPESCYLSSEDCGCHPOPLRAEMSAWARRCPPLSTEGELPEIHHRRMR 294  
 Db 240 NSDGNSEVVGQTLTPESSCLEDHCHPOPLSAELIPASWOCPCPLSTEGELPEIHHLRMK 299  
 QY 295 WLVEL-QPNQGDPLTDDQNGYHSLREHNLRLMDPOHCTDNPQAQVSPAADRP---E 349  
 Db 300 RLEFLQASKGQDLPTPDQDNGYHSLREHNLRLMDPKHCRDNPQTQFVPAAGDIPGNTQE 359  
 QY 350 PTEKPELVIQEV-----SQSPQSSSLFCFLPVEKECEEDHTNATDLSDRGESLPVSTRP 404  
 Db 360 STEEKIELLTVEVPLALEEFSFEGCPSEETPMEKEPCEGRISVVDYVSYLGEGLPISARP 419  
 QY 405 VCSNKLIDYILGGAPSDLEASDSESDWGEPEDDGDFSDGSLSESVDQSDSGLHLWN 464  
 Db 420 ACSNKLIDYILGGASSDLETSDPEGEDWDEAEEDGDFSDSLSDSLEQDPEGLHLWN 479  
 QY 465 SFHSVDPPKPNFTATIQTAARIAPRDSDSGTSWSSGCGV-GSCQEGPLPETDHSSE 523  
 Db 480 SFCSDVPNPNQFTATIQTAARIIVPEEPSDSEKOLSGKSDLENSQSGLSPETPEHSSGE 539  
 QY 524 EDDNEPSADAEENLKNMNSFCHSEDPYNLLNFKAPQPSGKNWKGRQDSKASSEVTVAFS 583  
 Db 540 EDDWESSADEAESLKNMNSFCHSDDPYNPLNFKAPFQTSGENEKGCRDSTKPSISVAIS 599  
 QY 584 GHHTLLSCAQLLESQENCPGCGLGEALAGERYTHIKRKKVTFLEEVTEYIISGDEDRK 643  
 Db 600 ECHTLLSCVKQLLGSQSECPDQVDRDVLSSGRHTRVKKKVTFLFEVTEYIISGDEDRK 659  
 QY 644 GPWEEFARDGCRFQKRIQETEVATGYCLAFEHREKMNRLR 684  
 Db 660 GPWEEFARDGCRFQKRIQETEDALGYCLTFEHRERMFNRLQ 700

RESULT 4

Q6PEGO

ID Q6PEGO PRELIMINARY; PRT; 408 AA.

AC Q6PEGO;

DT 05-JUL-2004 (TrEMBLrel. 27, Created)

DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)  
 DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)  
 DE PP1R15B protein.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6; TISSUE=Brain;  
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.2426038999;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Uedin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
 RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,  
 RA Jones S.J., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences".  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
 RN [2]  
 RC SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6; TISSUE=Brain;  
 RA Strausberg R.;  
 RL Submitted (Sep-2003) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC058078; AAH58078.1; --  
 DR GO; GO:0000164; C:protein phosphatase type 1 complex; IDA.  
 DR GO; GO:0004732; F:protein serine/threonine phosphatase activity; IDA.  
 DR GO; GO:0045182; F:translation regulator activity; IC.  
 DR GO; GO:0006983; P:ER-overload response; IDA.  
 DR GO; GO:0006446; P:regulation of translational initiation; IC.  
 DR GO; GO:0042542; P:response to hydrogen peroxide; IMP.  
 DR GO; GO:0006979; P:response to oxidative stress; IDA.  
 SQ SEQUENCE 408 AA; 45432 MW; 44AA70EFE25796C4 CRC64;  
 Query Match 50.6%; Score 1878.5; DB 2; Length 408;  
 Best Local Similarity 90.6%; Pred. No. 4e-107;  
 Matches 367; Conservative 7; Mismatches 30; Indels 1; Gaps 1;  
 QY 1 METGTHRAKRPGLGFWFLPFLRRSHACSSFFPPSSRQNPNGNSALPERRTRYWTKL 60  
 Db 1 METGTHRAKRPGLGFWFLPFLRRSHACSSFFPPSSRQNPNGNSALPERRTRYWTKL 60  
 QY 61 LSQALLPSLQKLLMSQLSGGLIPTWLDFAASYSALRASRGRESADPTVQKSLSY 120  
 Db 61 LSQALLPSLQKLLMSQLSGGLIPTWLDFAASYSALRALRGRESAAPTQKSLSS 120  
 QY 121 TAGLFAKTRVVSTIALARGCTPVAVLVLRLEVKLKAQERLDAAPFLLEQQLWGVEL 180  
 Db 121 LRLD-SSEDLVVSSLDWLEGLQWQSSSDLELKLKAQERLDAAPFLLEQQLWGVEL 179  
 QY 181 LPSLQAGLVSHRELDSSSGSLPSVQSLGNFKVSVYLLNPSYLDLPOLGLRCOSSAGGG 240  
 Db 180 LPSLQAGLVSHRELDSSSGSLPSVQSLGNFKVSVYLLNPSYLDLPOLGLRCOSSAGGG 239  
 QY 241 QFVGFRITLTPESCYLSSEDCGCHPOPLRAEMSAWARRCPPLSTEGELPEIHHRRMWLFQ 300  
 Db 240 QFVGFRITLTPESCYLSSEDCGCHPOPLRAEMSAWARRCPPLSTEGELPEIHLRMKRFELQ 299  
 QY 301 PNQGDPLTDDQNGYHSLREHNLRLMDPOHCTDNPQAQVSPAADRPPETEKPELVIQ 360  
 Db 300 ANKGELPTPDQNGYHSLREHNLRLMDPOHCTDNPQAQVSPAADRPPETEKPELVIQ 359





```
QY 280 LSTGLPEIHHRRMRWLVLQPNQGDLPDQDNGHSLSEENLHMDPQHCTDNPQAQ 339
Db 200 TPVPFLGEAEHQATE-----EKGTEKADPNSSPSGSHSRAWEY-YGREKPKQGEAKVE 254
QY 340 AVSPAADRP-----EPTKPKELVLEQVSSQSGSLFCELPVKECEDEHNTATDLSDRG 395
Db 255 AHRAGQHPCBNAAEAGGDETFP--VCTGNFLKAWVYRPGEDTEEDSDSDSAEDT 312
QY 396 ESLPVRTPVCSNKLIDYILGGAPSDLEASDSSESD-----W-----GE 435
Db 313 AQTGATPHTSAFLKAWVYRPGEDTEEDSDSDSAEDTAQTGATPHTSAFLKAWVYRPG 372
QY 436 EPEDDGDSDGLSSESDVEQDSEGLHLWNSFHSVDPYKPNQFTATIGTAARIAPRDPDS 495
Db 373 DTEENSDDLDS--ABEDTAQTGATPHT--SAFLKAWVYRPGEDTEEDSDSDSAEDTAQT 429
QY 496 GTSMGSCGVGSCQEGPLPETPDHSSGDEEDDWEPSADEAENLKLWNSFCHSEDPYNLNF 555
Db 430 GATPHTSPFLKAWVYRPGEDTEEDSDSDSAEDTADSSQSPC----- 478
QY 556 KAPFQPS-----GKNWKGRQDSKASSEVTVAFSGHH-----TLLSCKAQLLES-- 598
Db 479 --LQPCRLPGKTKGRGEPPLFQVAFVLPGEKPSPWAAPKLPLRLQRLRLPKAPT 535
QY 599 --QEDNCPGCGGLGALAGERYTHIKRKVTFLEBVTYYIS-----GDEDRKGWBEFAR 651
Db 536 RDQDPEIP-----LKKARKVHFAEKVTVHFLAVWAGPAQAARRGPEQFAR 580
QY 652 DGCRCFKRKIOETEAIVAGVCLAFEHREKMFNRLR 684
Db 581 DRSRFRARIAQAEKUGPYLTPDSRARAWAKLR 613

RESULT 9
Q6IN02 PRELIMINARY; PRT; 578 AA.
AC Q6IN02;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DE Mvdl16 protein.
GN Name=Mvdl16;
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUS=Heart;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strauberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.P., Zeeberg B., Buetow K.H., Schaefer C.P., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,
RA Boeak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettman M., Madan A.C., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grinstead J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Small D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUS=Heart;
```

```
RA Strauberg R.;
RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC072513; AAT72513.1; -.
SQ SEQUENCE 578 AA; 63583 MW; D9387CA71923345F CRC64;

Query Match 5.1%; Score 190.5; DB 2; Length 578;
Best Local Similarity 22.8%; Pred. No. 0.002;
Matches 131; Conservative 58; Mismatches 214; Indels 171; Gaps 30;

QY 216 YLLNPSYLDLPQLGLR-----COSSAGGQGVGFRTLTIPESCYLSBDGCHPQP 264
Db 20 YLLSP-LMGFLSRWSRLRGPEVSEAWLAETVAGANQIEADALLTTPP--VSENLH---P 73
QY 265 LR-----AEMSATAWRC-----PPLSTGLPEIHHRRMRWLVLQPNQGDLP 310
Db 74 LRETEGNGTPWMSKAAQRLCLDVEAQSSPPKTWGLSLDIDHNGK-----PQD--- 121
QY 311 DDNGYHSLSEEH--NLLRMDPOHC--TONPAQAV-----SPAADR 347
Db 122 ----GLREQVEHTAGLPTLQPLHQAADKKGVEGVAREEGVSELAYPTSHWEGGPADE 177
QY 348 PEPTKKPELVIOEVQSP---QGSLFCFELPVEKECED-----HTNATDLS 392
Db 178 EDTETVKKAQAASAAIAPGKPESTSVYCPGEAEHRAATEKGTDNKAEPGSGSHRWVEYH 237
QY 393 DR-----GESLPVSTRP---VCSNKLIDYILGGAPSDLEASDSSESDW-----GEEPE 438
Db 238 TRERPQKEGETKEQHRAGQSHPCNAEAE---EGGPETSVCSGSAFLKAWVYRPGEDTE 294
QY 439 DDGFDSDGSLSESDVEQDSEGLHLWNSFHSVDYKPNQFTATIGTAARIAPRDPDSGTS 498
Db 295 EEE--DSDLSAEDDTAHTCTTPT--SAFLKAWVYRPGEDT-----BEEDDG-D 339
QY 499 WSGSCGVGSCQEGPLPET-----PDHSSGDEEDDWEPSAD-EAENLKLWNSFCH-- 545
Db 340 WD-SAEEDTAQSCCTTHTSAFLKAWVYRPGEDTEEDSDSENAPVSETVDSQSTQCL 398
QY 546 -----SEDPYNLLNFKAPFPQSGKNWKGRQDSKASSEVTVAFSGHHTLLSCKAQL 595
Db 399 PVKTKGCGEAEPPFQVAFVLPQKPAPEW-----AAPKLPLRL--QKRLRSFKAP- 448
QY 596 LESQDNCPGCGGLGEALAGERYTHIKRKVTFLEBVTYYIS-----GDEDRKGWBEFA 650
Db 449 ARNQDPEIP-----LKRKVKHFSEKVTVHFLAVWAGPAQAARRGPEQFA 493
QY 651 RDGCRFKRKIOETEAIVAGVCLAFEHREKMFNRLR 684
Db 494 RDRSRFRARIAQAEKUGPYLTPAPRARAWTRLR 527

RESULT 10
Q7TQC2 PRELIMINARY; PRT; 578 AA.
AC Q7TQC2;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DE Protein phosphatase regulatory subunit 15A.
GN Name=Gadd34;
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley;
RX MEDLINE=22697643; PubMed=12813455; DOI=10.1038/sj.onc.1206567;
RA Hollander M.C., Poole-Kella S., Fornace A.J. Jr.;
RT "Gadd34 functional domains involved in growth suppression and
RT apoptosis."
RL Oncogene 22:3827-3832(2003).
DR EMBL; AY128642; AAM77795.1; -.
SQ SEQUENCE 578 AA; 63569 MW; 923BC49921C0BC61 CRC64;
```

```

Query Match          5.1%; Score 189.5; DB 2; Length 578;
Best Local Similarity 22.8%; Pred. No. 0.0023;
Matches 131; Conservative 58; Mismatches 214; Indels 171; Gaps 30;

QY 216 YLLNPSYLDYLPQLGLR-----CQSSAGGQGVFGFTLTPESCYLESDGCHPQP 264
D 20 YLLSP-LMGFLSRAWSRLRGPEVSEAWLAETVAGANQIEADALLTPPP--VSENLH--P 73

QY 265 LR-----AEMSAFAWRRC-----PPLSTEGLPETHHRRMRWLVLQPNQODLPTL 310
D 74 LRETEGNTPEWSKAAQRLCLDVEAQSPPKTWGLSDIDEHNGK-----PGQD---- 121

QY 311 DDQNGVHSLSEEH--NLRMDPQHC--TDNPAQAV-----SPAADR 347
D 122 ---GLRQEVHTAGLTQLQLHLOGADKKVGEVVARBEGVSELAYTSHWEGCPAED 177

QY 348 PEPTKKPVLVTQVQSP--QGSLSLFCFLPVEKCEED-----HTNATDLS 392
D 178 EDTETVKKAHQAASAASIAPGVKPTSVYCPGEAEHRAETEKGTDNKAEPGSGSHSRWEYH 237

QY 393 DR-----GESLPVSTRP-----VCSNKLIDYILGGAPSDLEASDSESDW----GEPE 438
D 238 TRERPKQGETKPEQHRAGQSHPCQNAEAE---EGGPETSVCSGSAFLKAWVYRPGEDTE 294

QY 439 DDGFGDSGLSESDVEQDSEGLHLWNSFHSVDYPKPNFTATQTAARIAPRDPDSGTS 498
D 295 EEE-DSLDLSAEDTAHTCTTPHT-SAPLKAWVYRPGEDT-----EEDDG-D 339

QY 499 WSGCGVSGCQGLPET-----PDHSSGEEDWEPESAD-EAENLKLWNSFCH-- 545
D 340 WD-SAEEDASQCTTPHTSAFLKAWVYRPGEDTEEDDSENVAPVDSETVDSQSTQHCL 398

QY 546 -----SEDVYNLNFKAPQPSGKNWKGQDSKASSEVTVAFSGHHTILSCKAQL 595
D 399 PVEKTKGCEAEPPPPQVAFYLPQCKPAPP-----AANKPLRL--QKRLSFKAP- 448

QY 596 LESQDNCPCGCLGALAGERYTHIKKKVTFLEEVTEYIS-----GDERRKGPWEFA 650
D 449 ARNQDPEP-----LKGKRVHSEKVTYHFLAVWAGPAQAARGFWEQFA 493

QY 651 RDGCRFQKRIQETVVAIGCYLAFEHREKMFNRLR 684
D 494 RDRSRFARRIAQAEQQLGPLYLTPAFRARAATRLR 527

RESULT 11
ID O75807 PRELIMINARY; PRT; 674 AA.
AC O75807;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Apoptosis associated protein (protein phosphatase 1, regulatory
DE subunit 15A).
GN Name=GADD34; Synonyms=PPP1R15A;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97298078; PubMed=9153226; DOI=10.1074/jbc.272.21.13731;
RA Hollander M.C., Zhan Q., Bae I., Fornace A.J. Jr.;
RA "Mammalian GADD34, an apoptosis- and DNA damage-inducible gene.";
RL J. Biol. Chem. 272:13731-13737(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,

```

```

RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Iaquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bobak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smallos D.E., Schnerch A., Schein J.E.,
RA "Generation and initial analysis of more than 15,000 full-length human
RA and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [3]
RC SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Strausberg R.;
RA Submitted (Feb-2001) to the EMBL/GenBank/DBJ databases.
RL EMBL; U83981; AAC25631.1; -.
DR EMBL; BC003067; AAH03067.1; -.
DR Genbank; HGNC:14375; PPIR15A.
DR GO; GO:0006915; P:apoptosis; TAS.
DR GO; GO:0007050; P:cell cycle arrest; TAS.
DR GO; GO:0006974; P:response to DNA damage stimulus; TAS.
DR GO; GO:0006974; P:response to DNA damage stimulus; TAS.
SQ SEQUENCE 674 AA; 73477 MW; B257AAL1456D1403 CRC64;

Query Match          5.1%; Score 189.5; DB 2; Length 674;
Best Local Similarity 21.5%; Pred. No. 0.0028;
Matches 156; Conservative 74; Mismatches 269; Indels 227; Gaps 32;

QY 60 LLSQLALLPSLFQKLLNSQLSGGLIPTR-WLDPAAYSALRASRGRESDAPTQVKS 118
D 21 LLSVPMGLLSR-----ANSRLR-GLGLEPFWLVEAVKGAAL----- 55

QY 119 SYTAAGLFAKTRVVTLLALARGGTPVAVLRLVLEVKLKAQERALDSAP-----T 168
D 56 --VEAGLEGAR-----TPLA--IPHTFWRRPPEEAEDSGPGEDRETGLKT 100

QY 169 FLLEQQLWGVELLPSLQAGLVSHRELDSSSGPLSVQSLGNFKVWSYLLNPSYL----- 223
D 101 SSSLPEAWG--LLDD--DDGMVGERATSVPRGQSQFADGQ-----RAPLSPLLIRTLQ 152

QY 224 --DYLQGLRQCSSAGGQGVFGFTLT-----PESCYLESDGCHPQPLAEMSAFA-- 273
D 153 GSDKNP--GEEKAEFGEVAGEEGVKNFSPYPPSHRECCPAVEEEDDEEA VKKEAHTRTSA 210

QY 274 -----WRRCP-----PLSTEGLPETHHRRMRWLVLQPNQODLPTLDQNGVHSL 320
D 211 LSPGSKPTWVSCPEENQATDKRTSKGARKTSVSPRSGSDPSRSEYRSGEASEE 270

QY 321 EHNLLRMDPQHCTDNPAQAVSPAADRPEPEKPELVIOEVQSQSPGSSSLFCFLPVEKE 380
D 271 KE-----EKAKHETGKEAA-----PGQSSAP-----AQRPLKQWMCQPSDEEE 311

QY 381 CEDHTNATDLSRGSPLVSTPRVCSNKLIDYILGGAPSDLEASDSESDWGEEREDD 440
D 312 GEVKALGAAEKDGEAECCPPCIPPPSAFLKAWVWPG---EDTDEEEDEE---DESDS 364

QY 441 GFSDGSLSDVEQDSEGLHL--WNSFHSVDYPKPNFTATIQ-----TAARIAPRDP 493
D 365 GSDEEAGEAEASSTPATGVFLKSW-----VYQGEDTEEEDESDSGSAEDEAE 417

QY 494 DSGTSMGSGCVSGCQGLPPTPDH-----SGBEDDWEPSADEAENLKLWNSFCHSE 547
D 418 TSASTPPASAFKAWVYRPGEDTEEEDEDDVDSEKDDSEAAALGEARS----- 466

QY 548 DPNYLLNFKAPFOPSG-----KNWKGQDSKASSEVTVAFSGHHTILSCKAQLLESQDNC 603
D 467 DPH-----PSHPDQRAHFRGWYRPGKETEE-----EAAED-- 498

```





Db 21 LLSVPMGLLSR-----TWSRLR-GLGFLPWLVEAVKGAAL----- 55  
QY 119 SYTAAGLFAKTRVSTLARGGTVAVLRLVLEKKAQERALDSAAP-----T 168  
Db 56 --VEAGLSGEAR-----TFLA--IHTPWGRPEAEADSGCGEDRETILGKT 100  
QY 169 FLLEQOLMGVLLPSSLAQGLVSHRELDSGSSGLPSVOSLGNFKVVSYLLNPSYL----- 223  
Db 101 SSSLPEAWG--LLDD--DDGMYGEREATSVPRGQSQFADGQ-----RAPLSFSLIRTLQ 152  
QY 224 --DYLPLQLGLRCQSSAGSQFVGFTLT-----PSCYLSGDCGCHPQLRAEMATA--- 273  
Db 153 GSDKNP--GEKAELEGVAREEGVKNKFSYPPSHRECCPAVEEDDEEAVKKEAHTTSA 210  
QY 274 -----WRCP-----PLSTEGLPETHHRRMRMLVFLQPNQGDLPFLDQDNGYHSLE 320  
Db 211 LSPGSKPTWVSCPEEENQATEDKRTSKGAKTYSVPSRSSGSDPRSWYRSGEASEE 270  
QY 321 BEHNLRLMDPOHCTDNPAQAVSPAADRPETKPKELVIOVVSQSPQSSSLFCPLPVEKE 380  
Db 271 KE-----EKAHKTGKEAA-----PGQSSAP-----AQLQLKSWWCQPSDEE 311  
QY 381 CEEDHTNATDLSDRGESLPVSTRPVCSNKLIDYILGAPSLDLEASSDSESDWGEPEDD 440  
Db 312 GEVKALGAARKDGEACPCIPPPSAFLKAWYTPG-----EDTEEEDEE-----DESDS 364  
QY 441 GFDGSLSESDVEQDSGLHL--WNSFHSVDYKPNFTATIO-----TAARTAPRDP 493  
Db 365 GSDEEGAEARASSTPATGVFLKSH-----VTPGEDTEEEDESDTGSADERAE 417  
QY 494 DSGTSWSCGCGVSCQEGPLPDPH-----SSGEEDDWEPSSADEAENLKLWNSFCHSE 547  
Db 418 TSASTPPASAFKAWYVRPGEDTEEEDEVDSEKEDSEALGEAES----- 466  
QY 548 DPNYLLNFKAPQPSG-----KNWGRQDSKASSEVTVAFSGHHTLLSCKAQLLESQEDNC 603  
Db 467 DPH-----PSHPDQRAHFRGMYRPGKETEE-----EAEED-- 498  
QY 604 PCGGLGEA-----LAGER-----YTHIKRK 623  
Db 499 -----WGEAEPCEFRVAIVYVEKEPPPPWAPRLPLRLQRLKRPTPHDPPETPLKAR 554  
QY 624 KVTFLVEVTEYIIS-----GDEDRKGPWEFARDGCFQKRIQIETVAIGYCLAPEHREK 678  
Db 555 KVPSEKVTVHFLAWAGPAQAARQGPWEQLARDRSFARRITQAEELSPLCPAARAR 614  
QY 679 MFNRLR 684  
Db 615 AWARLR 620

## RESULT 14

Q60465 PRELIMINARY; PRT; 590 AA.  
AC Q60465;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Gadd34 protein.  
GN Name=Gadd34;  
OS Cricetus longicaudatus (long-tailed hamster) (Chinese hamster).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;  
OC Cricetus.  
OX NCBI\_TaxID=10030;  
RN [1]  
RP MEDLINE=94187707; PubMed=8139541;  
RA Zhan Q., Lord K.A., Alamo I.Jr., Hollander M.C., Carrier F., Ron D.,  
RA Kohn K.W., Hoffman B., Liebermann D.A., Fornace A.Jr.;  
RT "The gadd and Myd genes define a novel set of mammalian genes encoding  
RT acidic proteins that synergistically suppress cell growth.";  
RL Mol. Cell. Biol. 14:2361-2371(1994).

DR EMBL; L28147; AAA36983.1; -.  
DR PIR; A56535; A56535.  
SQ SEQUENCE 590 AA; 64527 MW; B3D879BDACBAE6D6 CRC64;  
Query Match 4.6%; Score 171.5; DB 2; Length 590;  
Best Local Similarity 18.1%; Pred. No. 0.03; 198; Indels 187; Gaps 18;  
Matches 100; Conservative 68; Mismatches 198;  
QY 250 PESCYLSEDCGCHPQLRAEMASATAWRRCPPL-----STEGLPETHHRR 292  
Db 103 PETILGLSDDDKQGDGPREQGRAHTAGLPILLSGLQSDKSLGEVVAAGEGVTELAYPT 162  
QY 293 MRMLVFLQPNQGDLPFLDQDNG-----YHLSLEEH 323  
Db 163 SHW-----EGCSEEEDEGTVKAFRASADSPGHKSSTSVYCPGAEHQATEKQ 213  
QY 324 NLLRMDP-----QHCTDNPAQAVSPAADREP-----TEKKPELV 358  
Db 214 TENKADPPSPSGSHSRAWEYCSKQGEA-----DPEHRAKGYQLCQNAEAESEBEAK 267  
QY 359 IOEVSQSPQSSSL--FCLELPVEKECEDHTNATDLSDRGESLPVSTRPVCSNKLIDYILG 416  
Db 268 VSSLVSSGNAFLKAWYVRPGEDTDDDDSDWGAEEGKALSSPTSPE-----HDFLKA 322  
QY 417 GAPSDLEASSDSESDWGEPEDDGDFSDGSLSESDVEQDSGLHLMNSFHSVDYKPNQ 476  
Db 323 WYVRPGSDTDDDDSDWGSABE-----EGKALSSPTSPEHDFLKA-----VYRPG 369  
QY 477 FTATIQTAAARIAPRDSGTSWSCGCGVSCQEGPLPET----- 516  
Db 370 DT-----EDDQSDWGAEEKDGLAOTFATPHTSAFLKTVWCPCGEBTDDDC 416  
QY 517 ----PDHSSGEEDWEPSS-----ADEAENLKLWNSFCHSEDPYLLNFKAPQPSGK 564  
Db 417 EVVVPEDSEADPKSPSHEAQGLPCEQTEGL-----VEAHSFLFOVAYFLPGEKAP 470  
QY 565 NWKGRQDSKASSEVTVAFSGHHTLLSCKAQLLESQEDNCPGCGLGEALAGERYTHIKRK 624  
Db 471 PW-----TAPKLPLRLQRLRLTLRTPTQ---DQDPETP-----LRARK 505  
QY 625 VTFLVEVTEYIIS-----GDEDRKGPWEFARDGCFQKRIQIETVAIGYCLAPEHREK 679  
Db 506 VHFSENVTVHFLAWAGPAQAARQGPWEQLARDRSFARRIAQAEELKGLPYLTPAFARA 565  
QY 680 FNRLRTEKDLLL 692  
Db 566 WARLGNPSLPLAL 578

## RESULT 15

Q9VB94 PRELIMINARY; PRT; 676 AA.  
AC Q9VB94;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)  
DE CG6296-PA (LP07116p).  
GN ORENAMES-CG6296;  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Mortman J.R., Vandal M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,  
RA Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,



RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahke C., Davenport L.B., Davies P.,  
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K.J., Doup L.S., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacלב J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Slomson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.Y., Waserman D.A., Weinstein G.M., Weissenbach J.,  
RA Williams S.M., Woodgett, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,  
RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.; Smith H.O.;  
RT "Finishing a whole-genome shotgun: Release 3 of the Drosophila  
RT melanogaster euchromatic genome sequence.";  
RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).  
RN [3]  
RN SEQUENCE FROM N.A.  
RX MEDLINE=22426065; PubMed=12537568;  
RA Celniker S.E., Wheeler D.A., Krommiller B., Carlson J.W., Halpern A.,  
RA Patel S., Adams M., Champe M., Dugan S.P., Frisae E., Hodgson A.,  
RA George R.A., Hoskins R.A., Laverty T., Muzny D.M., Nelson C.R.,  
RA Pacלב J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,  
RA Svirkas R., Taber P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,  
RA Weinstein G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;  
RT "Finishing a whole-genome shotgun: Release 3 of the Drosophila  
RT melanogaster euchromatic genome sequence.";  
RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).  
RN [3]  
RN SEQUENCE FROM N.A.  
RX MEDLINE=22426070; PubMed=12537573;  
RA Kaminker J.S., Bergman C.M., Krommiller B., Carlson J., Svirkas R.,  
RA Patel S., Frisae E., Wheeler D.A., Lewis S.E., Rubin G.M.,  
RA Ashburner M., Celniker S.E.;  
RT "The transposable elements of the Drosophila melanogaster euchromatin:  
RT a genomics perspective.";  
RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
RN [4]  
RN SEQUENCE FROM N.A.  
RX MEDLINE=22426069; PubMed=12537572;  
RA Milera S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,  
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,  
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,  
RA Bettencourt B.R., Celniker S.E., de Grey A.D., Drysdale R.A.,  
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,  
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,  
RA Lewis S.E.;  
RT "Annotation of the Drosophila melanogaster euchromatic genome: a  
RT systematic review.";  
RL Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).  
RN [5]  
RN SEQUENCE FROM N.A.  
RG FlyBase;  
RL Submitted (SEP-2002) to the EMBL/GenBank/DBAJ databases.  
RN [6]  
RN SEQUENCE FROM N.A.  
RG FlyBase;  
RL Submitted (MAR-2004) to the EMBL/GenBank/DBAJ databases

Search completed: September 16, 2005, 10:46:02  
Job time : 75.2452 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 15, 2005, 08:08:59 ; Search time 1425.17 Seconds  
(without alignments)

12220.183 Million cell updates/sec

Title: US-10-650-482-1

Perfect score: 2942

Sequence: 1 attttggcttcgtccac.....aaattatgtgctctatctg 2942

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 100 summaries

Database :

N Geneseq\_16Dec04:\*

1: Geneseqn1980s:\*

2: Geneseqn1990s:\*

3: Geneseqn2000s:\*

4: Geneseqn2001as:\*

5: Geneseqn2001bs:\*

6: Geneseqn2002as:\*

7: Geneseqn2002bs:\*

8: Geneseqn2003as:\*

9: Geneseqn2003bs:\*

10: Geneseqn2003cs:\*

11: Geneseqn2003ds:\*

12: Geneseqn2004as:\*

13: Geneseqn2004bs:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2942	100.0	2942	12	ADQ88457
2	2934	99.7	2942	4	AAH18699
3	2185.8	74.3	2324	3	AAC98158
4	1324	45.0	5468	12	ADQ88459
5	828.6	28.2	883	4	AAH93389
6	738	25.1	772	4	AAH06623
7	515	17.5	543	4	AAH11666
8	265.8	9.0	561	6	ABL38451
9	218	7.4	218	3	AAC29803
10	199	6.8	706	6	ABK44494
11	198.6	6.8	399	8	ABX52167
12	195	6.6	195	3	AAC06938
13	187.4	6.4	628	6	AAH61819
14	162	5.5	173	10	ADD49398
15	61.8	2.1	2000	8	ADA71938
16	46.4	1.6	8346	6	ABK28328
17	45.8	1.6	15881	12	ADQ19901
18	44.2	1.5	193672	10	ADL13570
19	43.6	1.5	778	6	ABQ15588
20	43.6	1.5	778	6	ABQ15589

21	43.4	1.5	5986	6	ABK28397	Abk28397 DNA trans
22	43.2	1.5	6794	6	ABK31265	Abk31265 Signal tr
23	43.2	1.5	6794	6	ABL70220	AbL70220 Chemically
24	43.2	1.5	6794	6	AAH61175	AAH61175 Human gen
25	43.2	1.4	110000	12	ADH69807_0	ADH69807 Human vbe
26	42.2	1.4	267156	6	ABL68560	AbL68560 Kidney ca
27	42	1.4	2000	8	ADA71938	Ada71938 Rice gene
28	42	1.4	15649	4	AAH45397	AAH45397 Chemically
29	42	1.4	15649	6	ABK28242	Abk28242 DNA trans
30	41.6	1.4	1863	4	AAH16572	AAH16572 Human CDN
31	41.6	1.4	5983	10	ADA13381	Ada13381 Human int
32	41.6	1.4	32640	13	ACN37228	ACN37228 Human per
33	40.8	1.4	19236	6	ABN80227	ABN80227 Human che
34	40.6	1.4	629	13	ADQ51068	AdQ51068 Novel can
35	40.6	1.4	8093	8	ABZ10177	AbZ10177 Haematopo
36	40.6	1.4	8093	10	ADB54267	ADB54267 Pretreate
37	40.6	1.4	8093	10	ADH84189	ADH84189 Human lym
38	40.6	1.4	8093	13	ADH89565	ADH89565 Oligonuc
39	40.4	1.4	309	5	ABV61252	ABV61252 Human pro
40	40.4	1.4	6259	6	ABL32290	AbL32290 Human imm
41	40.2	1.4	461	5	ADL43330	AdL43330 Human ova
42	40.2	1.4	1293	13	ADS09901	ADs09901 Human the
43	40.2	1.4	1460	6	AAH62717	AAH62717 cDNA sequ
44	40.2	1.4	1560	6	ABX04170	ABx04170 Human mRN
45	40.2	1.4	1560	6	ABK84105	AbK84105 Human CDN
46	40.2	1.4	1560	6	ABN95913	ABn95913 Gene #241
47	40.2	1.4	1560	8	ABX55995	ABx55995 Gene enco
48	40.2	1.4	1560	10	ADC35146	ADc35146 Human bre
49	40.2	1.4	1560	11	ADL31851	ADl31851 Human CDN
50	40.2	1.4	1560	13	ACN39529	ACn39529 Tumour-as
51	40.2	1.4	1647	5	AAH80933	AAH80933 DNA enco
52	40.2	1.4	2410	5	ADL63044	ADl63044 Human ova
53	40.2	1.4	2908	12	ADQ29596	AdQ29596 Human col
54	40.2	1.4	3585	10	ADG32741	ADg32741 Human DNA
55	40.2	1.4	4252	12	ADH77169	ADe77169 Human CDN
56	40.2	1.4	6048	4	AAH46614	AAH46614 Tumour su
57	40.2	1.4	19345	6	ABN80017	ABn80017 Human che
58	40	1.4	752	6	ABQ54799	ABq54799 Human ova
59	40	1.4	6816	12	ADQ24856	ADQ24856 Human sof
60	40	1.4	7341	6	AAH61394	AAH61394 Human gen
61	40	1.4	18624	6	ABL33703	ABl33703 Human imm
62	40	1.4	61020	4	AAH46787	AAH46787 Tumour su
63	39.8	1.4	6167	6	ABK31407	ABk31407 Signal tr
64	39.8	1.4	6167	6	ABL70368	ABl70368 Chemically
65	39.8	1.4	6167	6	AAH61330	AAH61330 Human gen
66	39.8	1.4	6169	4	AAH46369	AAH46369 Tumour su
67	39.8	1.4	6169	6	ABN80096	ABn80096 Human che
68	39.8	1.4	16786	6	ABL34157	ABl34157 Human imm
69	39.6	1.3	20486	6	ABL34610	ABl34610 Human met
70	39.6	1.3	20486	7	ADS99871	ADs99871 Biophy
71	39.4	1.3	453	6	ABL56232	ABl56232 AmEPV apo
72	39.4	1.3	3523	8	ABZ10170	AbZ10170 Haematopo
73	39.4	1.3	6523	10	ADH84216	ADH84216 Human lym
74	39.4	1.3	6523	13	ADH89620	ADH89620 Oligonuc
75	39.4	1.3	50000	6	ABL56202	ABl56202 AmEPV gen
76	39.2	1.3	10047	8	ACF73323	ACf73323 Staphyloc
77	39.2	1.3	5311	6	ABL33019	ABl33019 Human imm
78	39.2	1.3	11147	8	ABZ10153	AbZ10153 Haematopo
79	39.2	1.3	14147	10	ADH54325	ADH54325 Pretreate
80	39.2	1.3	14147	10	ADH84163	ADH84163 Human lym
81	39.2	1.3	14147	13	ADH89523	ADH89523 Oligonuc
82	39.2	1.3	23439	2	AAV74349	AAv74349 Staphyloc
83	39.2	1.3	56153	4	AAH46794	AAH46794 Tumour su
84	39.2	1.3	61648	12	ADQ97663	ADQ97663 Human can
85	39	1.3	387	5	ADL71793	ADl71793 Human ova
86	39	1.3	387	5	ADL36946	ADl36946 Human ova
87	39	1.3	2990	6	ABQ54800	ABq54800 Human ova
88	39	1.3	8093	4	AAH46435	AAH46435 Tumour su
89	39	1.3	8093	6	ABK33973	ABk33973 Human DNA
90	39	1.3	8093	6	ABL92236	ABl92236 Chemically
91	39	1.3	8093	6	ABL49331	ABl49331 Human MLH
92	39	1.3	8093	8	ABZ10031	ABz10031 Haematopo
93	39	1.3	8093	8	ADA20360	ADA20360 Prostate

94 39 1.3 8093 8 ADA84167  
 95 39 1.3 8093 10 ADB54139  
 96 39 1.3 8093 10 ADE84113  
 97 39 1.3 8093 13 ADS89291  
 98 39 1.3 19087 6 ABL32792  
 c 99 38.8 1.3 757 2 AAX55996  
 100 38.8 1.3 6171 6 ABL33010

ADA84167 Human ren  
 Adb54139 Pretreate  
 Ade84113 Human lym  
 Ad889291 Oligonuc  
 Ab132792 Human imm  
 Aax55996 Human cDN  
 Ab133010 Human imm

# ALIGNMENTS

RESULT 1  
 ADQ88457  
 ID ADQ88457 standard; cDNA; 2942 BP.

XX AC ADQ88457;  
 XX 07-OCT-2004 (first entry)  
 XX Human GADD34-like (GADD34L) cDNA.  
 XX Oxidative stress; neuronal ischaemia; heart ischaemia; renal damage;  
 KW autoimmune disease; neurodegenerative disorder; therapy; GADD34-like;  
 KW GADD34L; eIF2alpha-specific regulatory subunit; phosphatase; human; gene;  
 KW ss.  
 XX Homo sapiens.

XX Key Location/Qualifiers  
 PH 407..2548  
 FT CDS /\*tag= a  
 FT /product= "Human GADD34-like (GADD34L) protein"  
 FT

XX US2004142345-A1.  
 PN 22-JUL-2004.  
 PD 28-AUG-2003; 2003US-00650482.  
 XX 06-SEP-2002; 2002US-0408679P.  
 PR (ROND/) RON D.  
 PA (JOUS/) JOUSSE C.  
 XX Ron D, Jousse C;  
 PI WPI; 2004-552556/53.  
 XX P-PSDB; ADQ88458.  
 DR GENBANK; AK027650.

XX Screening test substances for preventing or treating disease involving  
 PT oxidative stress, by testing test substances for its ability to inhibit  
 PT activity of GADD34L and identifying test substance that inhibits activity  
 PT of GADD34L.

XX Disclosure; SEQ ID NO 1; 30pp; English.  
 PS The present invention relates to a method of screening several test  
 CC substances for preventing or treating diseases involving oxidative stress  
 CC such as neuronal ischaemia, heart ischaemia, renal damage induced by  
 CC ischaemia or toxins, autoimmune diseases and neurodegenerative disorders.  
 CC The method involves testing the test substances for its ability to  
 CC inhibit the activity of GADD34-like (GADD34L), also referred to as  
 CC eIF2alpha-specific regulatory subunit of phosphatase, and identifying the  
 CC test substance which inhibits the activity of GADD34L. The present  
 CC sequence is human GADD34L cDNA.

SQ Sequence 2942 BP; 732 A; 739 C; 725 G; 746 T; 0 U; 0 Other;  
 Query Match 100.0%; Score 2942; DB 12; Length 2942;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 2942; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATTTTGGGCTTCGCTTCCACCGCACCGCGGCTTACCCAGTCCCTTCGGGTATCGCGTTG 60  
 DB 1 ATTTTGGGCTTCGCTTCCACCGCACCGCGGCTTACCCAGTCCCTTCGGGTATCGCGTTG 60  
 QY 61 CTGAGGGGCTTTTCAACCCCTCTGTAGTCGGAACACCATCGCCGAGGCGGTGGGGGAGCT 120  
 DB 61 CTGAGGGGCTTTTCAACCCCTCTGTAGTCGGAACACCATCGCCGAGGCGGTGGGGGAGCT 120  
 QY 121 CCTATCCATGGTGTGAAGCGTCGAGCGGACTAGGGAACCTCCTTCCCGCCAGATGGA 180  
 DB 121 CCTATCCATGGTGTGAAGCGTCGAGCGGACTAGGGAACCTCCTTCCCGCCAGATGGA 180  
 QY 181 AGTCGCATCAGTCGCGGCTATTTCGCGGGCTGTCTTCCCTGTGTTCTGCGGCCGCTG 240  
 DB 181 AGTCGCATCAGTCGCGGCTATTTCGCGGGCTGTCTTCCCTGTGTTCTGCGGCCGCTG 240  
 QY 241 CCGCATTCGCTGCCCTCTGTGGCTTTTCTGTGCTCGAAGATCGGCTCGAGCAGCGAC 300  
 DB 241 CCGCATTCGCTGCCCTCTGTGGCTTTTCTGTGCTCGAAGATCGGCTCGAGCAGCGAC 300  
 QY 301 GCCACCGCTGGGCGAGCGCGAGACTCTGTAGGCTTCTCCGAATCCCGTCGACCTCAGC 360  
 DB 301 GCCACCGCTGGGCGAGCGCGAGACTCTGTAGGCTTCTCCGAATCCCGTCGACCTCAGC 360  
 QY 361 CGCTGAGCGCGCGGCTTACCTGAGAGACTGTCAAGAAAGAGATGAGCGCGGAGC 420  
 DB 361 CGCTGAGCGCGCGGCTTACCTGAGAGACTGTCAAGAAAGAGATGAGCGCGGAGC 420  
 QY 421 AGCGGATCGCGGAACGGCTTGGCCCTCGGGGGGCTTCGGTTCGGGCACCCCTTTT 480  
 DB 421 AGCGGATCGCGGAACGGCTTGGCCCTCGGGGGGCTTCGGTTCGGGCACCCCTTTT 480  
 QY 481 CCCTCGGGGATCGCAAGCAGGCTTCTTAAGTTCCCGACGCTTCGCGCGGAAACTC 540  
 DB 481 CCCTCGGGGATCGCAAGCAGGCTTCTTAAGTTCCCGACGCTTCGCGCGGAAACTC 540  
 QY 541 CGGGAACCCACACTGCTTCTGCGCCAGCGCGAGACTCGGGTCAGTTACTGACGAA 600  
 DB 541 CGGGAACCCACACTGCTTCTGCGCCAGCGCGAGACTCGGGTCAGTTACTGACGAA 600  
 QY 601 ACTGCTCTCCAGCTCCTTGGCGGCTCCCGGATGCTTCAAGAGTGCTTAATTTGGAG 660  
 DB 601 ACTGCTCTCCAGCTCCTTGGCGGCTCCCGGATGCTTCAAGAGTGCTTAATTTGGAG 660  
 QY 661 CCACATTTTCGGTGGATGTTCCGACAGATGGCTAGATTTTGTCTGAGTCTACAGCGC 720  
 DB 661 CCACATTTTCGGTGGATGTTCCGACAGATGGCTAGATTTTGTCTGAGTCTACAGCGC 720  
 QY 721 CCTGAGAGCCTGGAAGGCGGGAACACAGCGCCGCCACAGCGCAGAGAACTTTTGGAG 780  
 DB 721 CCTGAGAGCCTGGAAGGCGGGAACACAGCGCCGCCACAGCGCAGAGAACTTTTGGAG 780  
 QY 781 TTGCTGTCAGCTCGACTCTTCAGACCCCTCGGTACACAGTCCCTTGAATGGCTAGAGGA 840  
 DB 781 TTGCTGTCAGCTCGACTCTTCAGACCCCTCGGTACACAGTCCCTTGAATGGCTAGAGGA 840  
 QY 841 GGGGATCCATGGGCAATACCTCGCCCGCAGACCTTAAATTTGGAGCTTAAGGCAAGGAAG 900  
 DB 841 GGGGATCCATGGGCAATACCTCGCCCGCAGACCTTAAATTTGGAGCTTAAGGCAAGGAAG 900  
 QY 901 TGCTTTTGAGCCCTGCGACACAGCGCTTTTCTTTAGAGCAGCAGCTGTGGGAGTGGAGCT 960  
 DB 901 TGCTTTTGAGCCCTGCGACACAGCGCTTTTCTTTAGAGCAGCAGCTGTGGGAGTGGAGCT 960  
 QY 961 GTTGCCAGTAGGCTTCAATCCCGTCTGTACTCTTAACCGGGAACTTGGCTCTTCGCCCTC 1020  
 DB 961 GTTGCCAGTAGGCTTCAATCCCGTCTGTACTCTTAACCGGGAACTTGGCTCTTCGCCCTC 1020  
 QY 1021 TGGGCTCTTAACATTCACGCGATAGCAATTTTCAGTGTGTATCTTATTTGCTGAACCC 1080  
 DB 1021 TGGGCTCTTAACATTCACGCGATAGCAATTTTCAGTGTGTATCTTATTTGCTGAACCC 1080

```
Qy 1081 TTCTACTGAGCTGCTTTCTTAGCTAGAAAGTCAGCTATCAGAACAGTGTGAAATAG 1140
Db 1081 TTCTACTGAGCTGCTTTCTTAGCTAGAAAGTCAGCTATCAGAACAGTGTGAAATAG 1140
Qy 1141 CGAGGTAGTCGGCTTCAGACACTAAACCCAGAGAGAGCTGCTTGAGAGGACCATTTG 1200
Db 1141 CGAGGTAGTCGGCTTCAGACACTAAACCCAGAGAGAGCTGCTTGAGAGGACCATTTG 1200
Qy 1201 TCATCCCCCAGCGCTGAGTGCAGAACTCATTTCCGGCCCTGTCGGCAGGAGATGTCACCTCT 1260
Db 1201 TCATCCCCCAGCGCTGAGTGCAGAACTCATTTCCGGCCCTGTCGGCAGGAGATGTCACCTCT 1260
Qy 1261 TTCTACGGAGGCTTACACAGAAATTCACCATCTTCGATGAAACGGCTGGAATTCCTTCA 1320
Db 1261 TTCTACGGAGGCTTACACAGAAATTCACCATCTTCGATGAAACGGCTGGAATTCCTTCA 1320
Qy 1321 ACAGGCTAACAGGGGGAAGATTTACCCACCCCTGACCAGGATATGGCTACCCAGCCT 1380
Db 1321 ACAGGCTAACAGGGGGAAGATTTACCCACCCCTGACCAGGATATGGCTACCCAGCCT 1380
Qy 1381 GGAGGAGGAACACAGCCTTCTCCGGATGGATCCAAAACA CTGCAGAGATACCCAAACACA 1440
Db 1381 GGAGGAGGAACACAGCCTTCTCCGGATGGATCCAAAACA CTGCAGAGATACCCAAACACA 1440
Qy 1441 GTTTGTTCTCTGCTGTGAGACATTCCTGGAACACCCAGGAATCCACTGAAGAAAAAT 1500
Db 1441 GTTTGTTCTCTGCTGTGAGACATTCCTGGAACACCCAGGAATCCACTGAAGAAAAAT 1500
Qy 1501 AGAATTATTACTACAGAGTTCCACTTGTCTTTGGAGAAGAGAGCCCTTCGAGGGCTG 1560
Db 1501 AGAATTATTACTACAGAGTTCCACTTGTCTTTGGAGAAGAGAGCCCTTCGAGGGCTG 1560
Qy 1561 TCCATCTAGTCAGATACCTATGGAAGAGAGCCTGGAGAGGCGCGAATAAGTGTAGTTGA 1620
Db 1561 TCCATCTAGTCAGATACCTATGGAAGAGAGCCTGGAGAGGCGCGAATAAGTGTAGTTGA 1620
Qy 1621 TTACTCATPACTAGAGGTGACCTTCCATTTCTGCCAGACAGCCTTGATGTAACAACT 1680
Db 1621 TTACTCATPACTAGAGGTGACCTTCCATTTCTGCCAGACAGCCTTGATGTAACAACT 1680
Qy 1681 GATAGATTATATTTGGAGGTGCATCCAGTGACCTGGAACAGTGTGATCCAGAGG 1740
Db 1681 GATAGATTATATTTGGAGGTGCATCCAGTGACCTGGAACAGTGTGATCCAGAGG 1740
Qy 1741 TGAGGATTGGATGAGGAAGCTGAGGATGATGTTTTCATAGTATAGTCACTGTGAGA 1800
Db 1741 TGAGGATTGGATGAGGAAGCTGAGGATGATGTTTTCATAGTATAGTCACTGTGAGA 1800
Qy 1801 CTCAGACCTTGAACAAGACCCCTGAAGGGCTTCACCTTTGGAACTCTTTCTGCAGTGTAGA 1860
Db 1801 CTCAGACCTTGAACAAGACCCCTGAAGGGCTTCACCTTTGGAACTCTTTCTGCAGTGTAGA 1860
Qy 1861 TCCTTATTAATCCCGAATCTTACAGCAACAAATTCAGACTGCTGCCAGAAATGTTCTTGA 1920
Db 1861 TCCTTATTAATCCCGAATCTTACAGCAACAAATTCAGACTGCTGCCAGAAATGTTCTTGA 1920
Qy 1921 AGAGCCTTCTGATTACAGAAAGGATTTCTCTGGCAAGTCGATCTAGAGAAATTCCTCCCA 1980
Db 1921 AGAGCCTTCTGATTACAGAAAGGATTTCTCTGGCAAGTCGATCTAGAGAAATTCCTCCCA 1980
Qy 1981 GTCTGGAAGCCTTCTGAGACCCCTGAGCATAGTTCTGGGAGGAGAGATGACTGGGAATC 2040
Db 1981 GTCTGGAAGCCTTCTGAGACCCCTGAGCATAGTTCTGGGAGGAGAGATGACTGGGAATC 2040
Qy 2041 TAGTCAGATGAGCAGAGAGTCTCAAACTGTGNACTCATTTCTGTAAATTCGTATGACCC 2100
Db 2041 TAGTCAGATGAGCAGAGAGTCTCAAACTGTGNACTCATTTCTGTAAATTCGTATGACCC 2100
Qy 2101 CTACAACCCCTTTAAATTTTAAGGCTCCTTTTCAAAACATCAGGGGAAAAATGAAAGGCTG 2160
Db 2101 CTACAACCCCTTTAAATTTTAAGGCTCCTTTTCAAAACATCAGGGGAAAAATGAAAGGCTG 2160
Qy 2161 TCGTGACTCAAGAGCCCCATCTGAGTCCATTTGTGGCCATTTTCTGAGTGTGCACACCTTACT 2220
```

```
Db 2161 TCSTGACTCAAGAGCCCCATCTGAGTCCAATGTGGCCATTTCTGAGTGTGCACACCTTACT 2220
Qy 2221 TTCTTTGAAGGTGCAGCTGTTGGGAGCCAAAGAAAGTGAATGTCCAGACTCGGTACAGCG 2280
Db 2221 TTCTTTGAAGGTGCAGCTGTTGGGAGCCAAAGAAAGTGAATGTCCAGACTCGGTACAGCG 2280
Qy 2281 TGAGCTTTCTTTCTGGAGGAAGACACACATGTCAAAAGAAAAAAGGTAAACCTTTCTTGA 2340
Db 2281 TGAGCTTTCTTTCTGGAGGAAGACACACATGTCAAAAGAAAAAAGGTAAACCTTTCTTGA 2340
Qy 2341 AGAAGTTACTGAGTATTTATATAGTGTGTATGAGGATCGCAAGACCATGGGAGAAATT 2400
Db 2341 AGAAGTTACTGAGTATTTATATAGTGTGTATGAGGATCGCAAGACCATGGGAGAAATT 2400
Qy 2401 TGCAAGGGATGGATGCAAGTTTCAGAAACGAAATTCAGAGAAACAGAAAGTCTATTGGATA 2460
Db 2401 TGCAAGGGATGGATGCAAGTTTCAGAAACGAAATTCAGAGAAACAGAAAGTCTATTGGATA 2460
Qy 2461 TTGCTTGACATTTTGAACACAGAGAAAGAAATGTTTAAATAGACTCCAGGGAACATGCTTCAA 2520
Db 2461 TTGCTTGACATTTTGAACACAGAGAGAAAGAAATGTTTAAATAGACTCCAGGGAACATGCTTCAA 2520
Qy 2521 AGGACTTAATGTTCTCAAGCAATGTTGAGTTGGCAGCCTGTAGTCTCTAGCTAGCATACAC 2580
Db 2521 AGGACTTAATGTTCTCAAGCAATGTTGAGTTGGCAGCCTGTAGTCTCTAGCTAGCATACAC 2580
Qy 2581 TACCTCTTACCTGAGAGGTGCTTTTTAAAAACAAATCTTGGCAGCTCTCTTTTGACATTT 2640
Db 2581 TACCTCTTACCTGAGAGGTGCTTTTTAAAAACAAATCTTGGCAGCTCTCTTTTGACATTT 2640
Qy 2641 TTTTCTTTAGAGGAAATGTAACTTGGATCTAGTTTAAATTTTGTGCAACATATCCC 2700
Db 2641 TTTTCTTTAGAGGAAATGTAACTTGGATCTAGTTTAAATTTTGTGCAACATATCCC 2700
Qy 2701 ACTCAGAAACATTCAGGTTTGAAGCCAGCCCTGATTAATGAAGATGAACCTAGTGTGATTT 2760
Db 2701 ACTCAGAAACATTCAGGTTTGAAGCCAGCCCTGATTAATGAAGATGAACCTAGTGTGATTT 2760
Qy 2761 CTAACTCTCCCTTTTGTAGTTTGTGATGCTTTTAAATGTCTTAAATGTCTGCATGAG 2820
Db 2761 CTAACTCTCCCTTTTGTAGTTTGTGATGCTTTTAAATGTCTTAAATGTCTGCATGAG 2820
Qy 2821 GTGAAAAGGGACCTTTTGTAGTTGTCATTTTGACATTTTCAAACTTTATTTCTTGAAA 2880
Db 2821 GTGAAAAGGGACCTTTTGTAGTTGTCATTTTGACATTTTCAAACTTTATTTCTTGAAA 2880
Qy 2881 ACAATATTTATAGGCTTAAAGCCCATTTTCATTTCTAAATCTAAATATGTGTGCTTATC 2940
Db 2881 ACAATATTTATAGGCTTAAAGCCCATTTTCATTTCTAAATCTAAATATGTGTGCTTATC 2940
Qy 2941 TG 2942
Db 2941 TG 2942
```

## RESULT 2

AAH18699

ID AAH18699 standard; cDNA; 2942 BP.

XX AC AAH18699;

XX AC AAH18699;

XX DT 26-JUN-2001 (first entry)

XX XX

XX DE Human cDNA sequence SEQ ID NO:18964.

XX XX

XX KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

XX OS Homo sapiens.

XX XX

XX XX

XX FN EP1074617-A2.

XX XX

XX PD 07-FEB-2001.

XX 28-JUL-2000; 2000EP-00116126.  
 XX 29-JUL-1999; 99JP-00248036.  
 XX 27-AUG-1999; 99JP-00300253.  
 PR 11-JAN-2000; 2000JP-00118776.  
 PR 02-MAY-2000; 2000JP-00183767.  
 PR 09-JUN-2000; 2000JP-00241899.  
 XX (HELI-) HELIX RES INST.  
 XX Ota T, Isogai T, Nishikawa T, Hayaashi K, Saito K, Yamamoto J;  
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
 XX WPI; 2001-318749/34.  
 XX  
 XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.  
 XX  
 XX Claim 8; SEQ ID NO 18964; 2537pp + Sequence Listing; English.  
 XX  
 XX The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention  
 XX  
 XX Sequence 2942 BP; 731 A; 741 C; 726 G; 744 T; 0 U; 0 Other;  
 SQ

Query Match 99.7%; Score 2934; DB 4; Length 2942;  
 Best Local Similarity 99.8%; Pred. No. 0;  
 Matches 2937; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ATTTTGGGCTTCCACCGCACAGCGGCTTACCGAGTCTTCGGGTATCGCGTTG 60  
 DB 1 ATTTTGGGCTTCCACCGCACAGCGGCTTACCGAGTCTTCGGGTATCGCGTTG 60

QY 61 CTCAGGGGCTTTCAACCTCTGTCAGTGGGAAACCTCCCGAGGCGGTGGGGGACT 120  
 DB 61 CTCAGGGGCTTTTCAACCTCTGTGTCAGTGGGAAACCTCCCGAGGCGGTGGGGGACT 120

QY 121 CCTATCCATGTGTGTAAGAGCTCGAGCGGACTAGGGAACTCTTCCCGCCGAGGATGGA 180  
 DB 121 CCTATCCATGTGTGTAAGAGCTCGAGCGGACTAGGGAACTCTTCCCGCCGAGGATGGA 180

QY 181 AGTGCATCAGTCCGGCTATTCGGGGCTGTTCCTCTGTGTTCTCGCGCCCGCTG 240  
 DB 181 AGTGCATCAGTCCGGCTATTCGGGGCTGTTCCTCTGTGTTCTCGCGCCCGCTG 240

QY 241 CGCATTCGTCGCTCTGTGGCTTTTCTGCTGGCTCGAAGATGGGCTGGAGCAGCGAC 300  
 DB 241 CGCATTCGTCGCTCTGTGGCTTTTCTGCTGGCTCGAAGATGGGCTGGAGCAGCGAC 300

QY 301 GCCACCGCTGGGCAAGGCGGAGACTCTGTAGGCTTCTCCGAATCCCGTTCGACCTCCAGC 360

DB GCCACCGCTGGGCAAGGCGGAGACTCTGTAGGCTTCTCCGAATCCCGTTCGACCTCCAGC 360  
 QY CGCTGAGGCGCGGCGCTTACCTGAGAGACTGTCAAGAAAAAGAGATGGAGCGGGAC 420  
 DB CGCTGAGGCGCGGCGCTTACCTGAGAGACTGTCAAGAAAAAGAGATGGAGCGGGAC 420  
 QY AGCGGATCGCGGAAACGGCTTGGCCCTCGGGGGGCTTCGGGTTCTGGCCACCTTTT 480  
 DB AGCGGATCGCGGAAACGGCTTGGCCCTCGGGGGGCTTCGGGTTCTGGCCACCTTTT 480  
 QY CCTCTCGGCGATCGCAAGCAGGCTTCTTAAGTTTCCCGACGCTCTTGGCCCGGAAAACTC 540  
 DB CCTCTCGGCGATCGCAAGCAGGCTTCTTAAGTTTCCCGACGCTCTTGGCCCGGAAAACTC 540  
 QY CGGAAACCCACACTGCTTCTGCGCCAGCCGAGACTCGGGTCAAGTACTGGAGCAA 600  
 DB CGGAAACCCACACTGCTTCTGCGCCAGCCGAGACTCGGGTCAAGTACTGGAGCAA 600  
 QY ACTGCTCTCCAGCTCTTGGCGGCTCCCGGATTCCTCAGAAAGTGTCTAATTGGAG 660  
 DB ACTGCTCTCCAGCTCTTGGCGGCTCCCGGATTCCTCAGAAAGTGTCTAATTGGAG 660  
 QY CCAACTTTTTCGGTGGAAATGTTTCCGACAGATGGCTAGATTTTGTCTGGAGTCTACAGCG 720  
 DB CCAACTTTTTCGGTGGAAATGTTTCCGACAGATGGCTAGATTTTGTCTGGAGTCTACAGCG 720  
 QY CTTGAGAGCTTGAAGGAGCGGAGAAACAGCGCCGCCCCCAGCGCAGAGAAATTTTGG 780  
 DB CTTGAGAGCTTGAAGGAGCGGAGAAACAGCGCCGCCCCCAGCGCAGAGAAATTTTGG 780  
 QY TTGCTCAGCTCGACTCCTCAGACCTCGTCCAGCTCCCTTCAATGGCTTGGCTAGAGGA 840  
 DB TTGCTCAGCTCGACTCCTCAGACCTCGTCCAGCTCCCTTCAATGGCTTGGCTAGAGGA 840  
 QY GGGGATCCACTGGCAATFACCTGCGCCCGCAGACCTAAAAATTCGAGCTTAAGGCCAAGGAAG 900  
 DB GGGGATCCACTGGCAATFACCTGCGCCCGCAGACCTAAAAATTCGAGCTTAAGGCCAAGGAAG 900  
 QY TGCTTTGGAGCTGCAGCAGAGCTTTTCTTTAGAGCAGCAGCTGTGGGAGTGGAGCT 960  
 DB TGCTTTGGAGCTGCAGCAGAGCTTTTCTTTAGAGCAGCAGCTGTGGGAGTGGAGCT 960  
 QY GTTCCCGAGTAGCTTCAATCCCGTCTGTACTCTAACCAGGAACTTGGCTTTCGCCCTC 1020  
 DB GTTCCCGAGTAGCTTCAATCCCGTCTGTACTCTAACCAGGAACTTGGCTTTCGCCCTC 1020  
 QY TGGGCTCTAAACATTCACGCAATAGCAATTTTCAGTGTGGTATCTTATTTGCTGAACCC 1080  
 DB TGGGCTCTAAACATTCACGCAATAGCAATTTTCAGTGTGGTATCTTATTTGCTGAACCC 1080  
 QY TTCTTACCTGAGCTGCTTTCTTAGGCTAGAGTCAAGTATCAGAACAGTATGGAATAG 1140  
 DB TTCTTACCTGAGCTGCTTTCTTAGGCTAGAGTCAAGTATCAGAACAGTATGGAATAG 1140  
 QY CAGGTAGTTCGGCTTCAGACACTAACCCAGAGAGCAGCTGCCTGAGAGAGGACCAATTG 1200  
 DB CAGGTAGTTCGGCTTCAGACACTAACCCAGAGAGCAGCTGCCTGAGAGAGGACCAATTG 1200  
 QY TCATTCGCGAGCTGAGTCAAGAACTCATTTCCGGCTCTGTGGCAGGATGTCCACTCT 1260  
 DB TCATTCGCGAGCTGAGTCAAGAACTCATTTCCGGCTCTGTGGCAGGATGTCCACTCT 1260  
 QY TTCTTACGGAAGGCTTACCAAAATTCACCATCTTCGATCAAAAGGCTGGAATTCCTTCA 1320  
 DB TTCTTACGGAAGGCTTACCAAAATTCACCATCTTCGATCAAAAGGCTGGAATTCCTTCA 1320  
 QY ACAGGCTTAAACAGGGCAAGATTTTACCCAGCTTCAGCAGGATAATGGCTACCAAGCCT 1380  
 DB ACAGGCTTACCAAGGGCAAGATTTTACCCAGCTTCAGCAGGATAATGGCTACCAAGCCT 1380  
 QY GGAGGAGGAAACAGAGCTTCTCGGATGGATCAAAACACTGCGAGATTAACCCAAACA 1440  
 DB GGAGGAGGAAACAGAGCTTCTCGGATGGATCAAAACACTGCGAGATTAACCCAAACA 1440

[illegible]

Qy	2521	AGGACTTAATGTTCTCAAGCAATGTTGAGTTGGCAGCGCTGAGTCCCTAGCTAGCATACAC	2580
Db	2521	AGGACTTAATGTTCTCAAGCAATGTTGAGTTGGCAGCGCTGAGTCCCTAGCTAGCATACAC	2580
Qy	2581	TACCTCTTTACCTGAGAGGGTGCTTTTAAAAACAAATCTTGGCAGCTGTCTTTTGACATTT	2640
Db	2581	TACCTCTTTACCTGAGAGGGTGCTTTTAAAAACAAATCTTGGCAGCTGTCTTTTGACATTT	2640
Qy	2641	TTTTTTTTTAGAGGAAATGTAACTTGGATCTAGTTTAAATTTTTTTTTTGGCAACATATCCC	2700
Db	2641	TTTTTTTTTAGAGGAAATGTAACTTGGATCTAGTTTAAATTTTTTTTTTGGCAACATATCCC	2700
Qy	2701	ACTCAGAAACATTCAGGCTTTGAAGCCGAGCCCTGATATGAAGATGAACATAGTGTGATTT	2760
Db	2701	ACTCAGAAACATTCAGGCTTTGAAGCCGAGCCCTGATATGAAGATGAACATAGTGTGATTT	2760
Qy	2761	CTAATCCTCCCTTTTTTTGATTTAGTTGGATGTGCTTTTAAATGTCTTTTGCCTGCGATGAG	2820
Db	2761	CTAATCCTCCCTTTTTTTGATTTAGTTGGATGTGCTTTTAAATGTCTTTTGCCTGCGATGAG	2820
Qy	2821	GTGGAAGGGGACCTTTTTCAGTTGTCAATTTTGCACCTTTCAAACTTATTTTCTTGGAAA	2880
Db	2821	GTGGAAGGGGACCTTTTTCAGTTGTCAATTTTGCACCTTTCAAACTTATTTTCTTGGAAA	2880
Qy	2881	ACAATATTTATAGGCGCTTAAAGCCCATTTTCATTTCTAAATCTAAATATGTGTGCGCTATC	2940
Db	2881	ACAATATTTATAGGCGCTTAAAGCCCATTTTCATTTCTAAATCTAAATATGTGTGCGCTATC	2940
Qy	2941	TG 2942	
Db	2941	TG 2942	
RESULT 3			
AAC98158			
ID	AAC98158 standard; cDNA; 2324 BP.		
AC	AAC98158;		
DT	09-MAR-2001 (first entry)		
DE	Human colon cancer antigen nucleotide sequence SEQ ID NO:168.		
KW	Human; colon cancer; colon cancer antigen; diagnosis; detection;		
KW	identification; cytostatic; cardioactive; neuroprotective; vulnerary;		
KW	immunomodulatory; muscular; gynaecological; gastrointestinal;		
KW	nephrotropic; antiinfective; antibacterial; gene therapy; wound;		
KW	neural disorder; immune system disorder; muscular disorder;		
KW	reproductive disorder; gastrointestinal disorder; renal disorder;		
KW	infectious disease; cardiovascular disorder; ss.		
OS	Homo sapiens.		
PN	WO200055351-A1.		
XX	21-SEP-2000.		
XX	08-MAR-2000; 2000WO-US005883.		
XX	12-MAR-1999; 99US-0124270P.		
XX	(HUMA-) HUMAN GENOME SCI INC.		
XX	Rosen CA, Ruben SM;		
XX	WPI; 2000-587534/55.		
XX	P-PSDB; RAB53401.		
PT	Colon cancer associated gene sequences, referred to as colon cancer		
PT	antigens, useful for the treatment, prevention, and diagnosis of colon		
PT	disorders such as colon cancer.		



Claim 1; Page 595; 2104pp; English.

PS AAC97991 to AAC98763 encode the human colon cancer associated proteins,  
XX called human colon cancer antigens, given in AB53234 to AB54006. The  
CC human colon cancer antigens can have cytotostatic, cardioactive, muscular;  
CC neuroprotective, immunomodulatory, gynaecological, gastrointestinal, and  
CC vulnary, nephrotropic, antineoplastic and antibacterial activities, and  
CC can be used in gene therapy. The colon cancer antigen polynucleotides,  
CC proteins and antibodies to the proteins are useful for the prevention,  
CC treatment and diagnosis of colon disorders, such as colon cancer. The  
CC polynucleotides may be used in diagnostics and research, such as for  
CC chromosome identification, and as hybridisation probes. The proteins may  
CC also be used to prevent diseases such as neural disorders, immune system  
CC disorders, muscular disorders, reproductive disorders, gastrointestinal  
CC disorders, wounds, renal disorders, infectious diseases, and  
CC cardiovascular disorders. AAC98764 to AAC98772 and AB54007 represent  
CC sequences used in the exemplification of the present invention  
XX  
SQ Sequence 2324 BP; 615 A; 566 C; 562 G; 574 T; 0 U; 7 Other;

Query Match 74.3%; Score 2185.8; DB 3; Length 2324;  
Best Local Similarity 97.9%; Pred. No. 0;  
Matches 2249; Conservative 6; Mismatches 36; Indels 6; Gaps 4;

QY 456 GCTTCGGGTTCTGGCCACCCCTTTTCCCTCGCGCATCGCAAGCAGGCTCTTCTAAGTTCC 515  
DB 32 GCTCAGGGGCTTTTCAACCCCTCTGTCAGTCGGCGATCGCAAGCAGGCTCTTCTAAGTTCC 91  
QY 516 CGACGCCCTCTTGGCCCGGAAACCTCGGGAAACCCACACCTGCTTTCCTCTGCGCAGCCCG 575  
DB 92 CGACGCCCTCTTGGCCCGGAAACCTCGGGAAACCCACACCTGCTTTCCTCTGCGCAGCCCG 151  
QY 576 AGACTCGGGTCAGTTACTGCGAAGCTCTCTCCAGCTCTTGGCGCGCTCCCGGAT 635  
DB 152 AGACTCGGGTCAGTTACTGCGAAGCTCTCTCCAGCTCTTGGCGCGCTCCCGGAT 211  
QY 636 TGCTTCAGAAAGTGTCTAATTTTGGAGCCAACTTTTCGGTGGAAATGTTTCCGACCAAGATGCG 695  
DB 212 TGCTTCAGAAAGTGTCTAATTTTGGAGCCAACTTTTCGGTGGAAATGTTTCCGACCAAGATGCG 271  
QY 696 TAGATTTTGTGGAGTCTACAGCGCCCTGAGAGCCCTGAGGAGCGGAGAGAAACAGCCCG 755  
DB 272 TAGATTTTGTGGAGTCTACAGCGCCCTGAGAGCCCTGAGGAGCGGAGAGAAACAGCCCG 331  
QY 756 CCCACACAGCGCAGAAATCTTTAGTTCGTCGACCTCGACTCTCCAGACCCCTCGGTCA 815  
DB 332 CCCACACAGCGCAGAAATCTTTAGTTCGTCGACCTCGACTCTCCAGACCCCTCGGTCA 391  
QY 816 CCAGTCCCTTGAATTTGGCTAGAGGAGGGATCCACTGGCAATPACTCGCCCCCAGACCTTAA 875  
DB 392 CCAGTCCCTTGAATTTGGCTAGAGGAGGGATCCACTGGCAATPACTCGCCCCCAGACCTTAA 451  
QY 876 AATTGGAGCTTAAAGCCAAAGGAAGTGTCTTGGAGCCCTGAGAGCCCTGAGAGCAGGCTTTCTTTAG 935  
DB 452 AATTGGAGCTTAAAGCCAAAGGAAGTGTCTTGGAGCCCTGAGAGCAGGCTTTCTCTTAG 511  
QY 936 AGCAGCAGCTGTGGGAGTGGAGCTGTGGCCAGTAGCTCTCAATCCCGCTGTGACTCTA 995  
DB 512 AGCAGCAGCTGTGGGAGTGGAGCTGTGGCCAGTAGCTCTCAATCCCGCTGTGACTCTA 571  
QY 996 ACCGGGAATCTTGGCTCTTCGCCCTCTGGGCCCTCTAAACAATCAACGCATAGACAAATTTCA 1055  
DB 572 ACCGGGAATCTTGGCTCTTCGCCCTCTGGGCTCTTAAACAATCAACGCATAGACAAATTTCA 631  
QY 1056 GTGTGGTATCTTATTTGCTGAACCCCTTCTTACCTGGACTGTCTTCTTAGGCTTAGAGTCA 1115  
DB 632 GTGTGGTATCTTATTTGCTGAACCCCTTCTTACCTGGACTGTCTTCTTAGGCTTAGAGTCA 691  
QY 1116 GCTATCAGAACAGTCAATAGCGAATAGCGAGTGTGGCTTCCAGACACTAAACCCAGAGA 1175  
DB 692 GCTATCAGAACAGTCAATAGCGAATAGCGAGTGTGGCTTCCAGACACTAAACCCAGAGA 751  
QY 1176 GCAGCTGCCTGAGAGAGGACCAATTTGTCTATCCCGCGCTGAGTGCAGAACTCATTTCCGG 1235

DB 752 GCAGCTGCCTGAGAGAGGACCAATTTGTCTATCCAGCCGCTGARTGCAGAACTCATTCGG 811  
QY 1236 CTTGCTGGCAGGATGTCACCTCTTTCTACGGAAGSCCTTACCAAGAAATTCACCATCTTC 1295  
DB 812 SCTGCTGGCAGGATGTCACCTCTTTCTACGGAAGSCCTTACCAAGAAATTCACCATCTTC 871  
QY 1296 GCATGAACCGCTGGAAATCTTTCAACAGGCTAACAGGGGCAAGATTATACCAACCCCTG 1355  
DB 872 GCATGAACCGCTGGAAATCTTTCAACAGGCTAGCAAGGGGCAAGATWTATCCCAACCCCTG 931  
QY 1356 ACCAGGAATATGGCTACCAAGGCTGAGGAGGAAACACAGCTTCTCCGATGATGATCAA 1415  
DB 932 ACCAGGAATATGGCTACCAAGGCTGAGGAGGAAACACAGCTTCTCCGATGATGATCAA 991  
QY 1416 AACACTGCAGAGATAACCCCAACAGTTTCTCTGCTGCTGGAGACATTCCTCGGAACA 1475  
DB 992 AACACTGCAGAGATAACCCCAACAGTTTCTCTGCTGCTGGAGACATTCCTCGGAACA 1051  
QY 1476 CCCAGGAATCCACTGAAGAAAAATAGAAATTAATACTACAGAGGTTTCCACTTGTCTTGG 1535  
DB 1052 CCCAGGAATCCACTGAAGAAAAATAGAAATTAATACTACAGAGGTTTCCACTTGTCTTGG 1111  
QY 1536 AAGAAGAGAGCCCTTCTGAGGGCTGTCCATCTAGTGAGATACTATGGAAGAGAGCCTG 1595  
DB 1112 AAGAAGAGAGCCCTTCTGAGGGCTGTCCATCTAGTGAGATACTATGGAAGAGAGCCTG 1171  
QY 1596 GAGAGGCCCAATTAAGTGTAGTTTACTCATACCTTAGAAGGTGACCTTCCCAATTTCTG 1655  
DB 1172 GAGAGGCCCAATTAAGTGTAGTTTACTCATACCTTAGAAGGTGACCTTCCCAATTTCTG 1231  
QY 1656 CCAGACCAAGCTTGTAGTAAACAACTGATAGATTATATTTTGGGAGGTGCAATCCAGTGACC 1715  
DB 1232 CCAGACCAAGCTTGTAGTAAACAACTGATAGATTATATTTTGGGAGGTGCAATCCAGTGACC 1291  
QY 1716 TGGAAAAAGTCTGATCCAGAAAGGTGAGGATTTGGGATGAGGAAGCTGAGGATGATGGTT 1775  
DB 1292 TGGAAAAAGTCTGATCCAGAAAGGTGAGGATTTGGGATGAGGAAGCTGAGGATGATGGTT 1351  
QY 1776 TTGATAGTGTAGTCTACTCTCAGACTCAGACTTGAACCAAGACCCCTGAAGGGCTTCACC 1835  
DB 1352 TTGATAGTGTAGTCTACTCTCAGACTCAGACTTGAACCAAGACCCCTGAAGGGCTTCACC 1411  
QY 1836 TTTGGAACTCTTTCTGCAAGTGTAGATCTTTAATATCCCCAGAACTTTTACAGCAACAATTC 1895  
DB 1412 TTTGGAACTCTTTCTGCAAGTGTAGATCTTTAATATCCCCAGAACTTTTACAGCAACAATTC 1471  
QY 1896 AGACTGTGCGCAGAAATTTGTTCTGAGAGAGCTTCTGATTCAGAGAGGATTTGTCTGGCA 1955  
DB 1472 AGACTGTGCGCAGAAATTTGTTCTGAGAGAGCTTCTGATTCAGAGAGGATTTGTCTGGCA 1531  
QY 1956 AGTCTGATCTAGAGAAATTTCTCCAGTCTGGAAGCCCTTCTGAGAGCCCTGAGCATAGTT 2015  
DB 1532 AGTCTGATCTAGAGAAATTTCTCCAGTCTGGAAGCCCTTCTGAGAGCCCTGAGCATAGTT 1591  
QY 2016 CTGGGAGGAGAGTGAAGTGGGAAATCTAGTCAGATGAAGCAGAGAGTCTCAAACTGTGGA 2075  
DB 1592 CTGGGAGGAGAGTGAAGTGGGAAATCTAGTCAGATGAAGCAGAGAGTCTCAAACTGTGGA 1651  
QY 2076 ACTCATTTCTGTAATTCGTATGAGCCCTTCAACCCCTTTAAATTTTAAAGGCTCTTTTCAA 2135  
DB 1652 ACTCATTTCTGTAATTCGTATGAGCCCTTCAACCCCTTTAAATTTTAAAGGCTCTTTTCAA 1711  
QY 2136 CATCAGGGGAAAAATGAGAAAGGCTGTGCTGACTCAAAAGACCCCACTGTGAGTCCAATTTGG 2195  
DB 1712 CATCAGGGGAAAAATGAGAAAGGCTGTGCTGACTCAAAAGACCCCACTGTGAGTCCAATTTGG 1771  
QY 2196 CCAATTTCTGAGTGTCAACCTTACTTTTGTAAAGTGTGAGTGTGGGAGGAGCAAGAA 2255  
DB 1772 CCAATTTCTGAGTGTCAACCTTACTTTTGTAAAGTGTGAGTGTGGGAGGAGCAAGAA 1831  
QY 2256 GTGATGTCCAGACTCGGTACAGGGTGAACCTTTCTTCTGGGAGGAGACACACATCTCA 2315

Db 1832 GTGAATGTCAGACTCGGTAGCGGTGACGTTCTTTCTGGAGGAGACACACATGTCA 1891

Qy 2316 AAAGAAAAAGGTAACTTCTCTTGAAGAAATTAAGTATATATAGTGGTATGAGG 2375

Db 1892 AAAGAAAAAGGTAACTTCTCTTGAAGAAATTAAGTATATATAGTGGTATGAGG 1951

Qy 2376 ATCGCAAGGACCAATGGGAAGAAATTTGCAAGGATGGATGAGTTCAGAAAACGAATTC 2435

Db 1952 ATCGCAAGGACCAATGGGAAGAAATTTGCAAGGATGGATGAGTTCAGAAAACGAATTC 2011

Qy 2436 AAGAAACAGAGATGCTATTCGATATTCCTTGACATTTGAACACAGAGAAAGATGTTA 2495

Db 2012 AAGAAACAGAGATGCTATTCGATATTCCTTGACATTTGAACACAGAGAAAGATGTTA 2071

Qy 2496 ATAGACTCCAGGGAACATGCTTCAAGGACATTAATGTTCTCAAGCAATGTTGAGTTGGCA 2555

Db 2072 ATAGACTCCAGGGAACATGCTTCAAGGACATTAATGTTCTCAAGCAATGTTGAGTTGGCA 2131

Qy 2556 GCCTGTAGTCTAGCTAGCATACACTCTTACCTGAGAGGTGTCTTTTAAACAA 2615

Db 2132 GCCTGTAGTCTAGCTAGCATACACTCTTACCTGAGAGGTGTCTTTTAAACAA 2191

Qy 2616 TCTTGGCAGCTGCTCTTGGACATTTTCTTTTATAGAGGAATGTAACTTGCATCTAGTTT 2675

Db 2192 TCTTGGCAGCTGCTCTTGGACATTTTCTTTTATAGAGGAATGTAACT--GGMCTGTAA 2247

Qy 2676 AATTTTTTTTTTGCACATATCCCACTCAGAAACATT-CAGGTTTGAAGCCAGCCCT-G 2733

Db 2248 TTTTTTTTTTGCACATATCCCACTCAGAAACATTCCAGGTTTGAAGCCAGCCCTGG 2307

Qy 2734 ATAATGAAGATGAAT 2750

Db 2308 ATAATGAAGATGAAT 2324

RESULT 4

ID ADQ88459 standard; cDNA; 5468 BP.

XX ADQ88459;

XX 07-OCT-2004 (first entry)

XX Mouse GADD34-like (GADD34L) cDNA.

XX Oxidative stress; neuronal ischaemia; heart ischaemia; renal damage; autoimmune disease; neurodegenerative disorder; therapy; GADD34-like; GADD34L; eIF2alpha-specific regulatory subunit; phosphatase; mouse; gene; ss.

XX Mus sp.

XX Key Location/Qualifiers

XX CDS 462..2558

XX /tag= a

XX /product= "Mouse GADD34-like (GADD34L) protein"

XX /transl\_except= (pos:1332..1334, aa:Arg)

XX /transl\_except= (pos:1341..1343, aa:Arg)

XX /transl\_except= (pos:1344..1346, aa:Trp)

XX /transl\_except= (pos:1350..1352, aa:Val)

XX /transl\_except= (pos:1362..1364, aa:Pro)

XX /transl\_except= (pos:1368..1370, aa:Gln)

XX /transl\_except= (pos:1377..1379, aa:Asp)

XX /transl\_except= (pos:1389..1391, aa:Leu)

XX US2004142345-A1.

XX 22-JUL-2004.

XX 28-AUG-2003; 2003US-00650482.

XX 06-SEP-2002; 2002US-0408679P.

XX

PA (ROND/) RON D.

PA (JOUS/) JOUSSE C.

PI Ron D, Jousse C;

DR WPI; 2004-552556/53.

DR P-PSDB; ADQ88460.

PT Screening test substances for preventing or treating disease involving oxidative stress, by testing test substances for its ability to inhibit activity of GADD34L and identifying test substance that inhibits activity of GADD34L.

PT

XX Disclosure; SEQ ID NO 3; 30pp; English.

XX The present invention relates to a method of screening several test substances for preventing or treating diseases involving oxidative stress such as neuronal ischaemia, heart ischaemia, renal damage induced by ischaemia or toxins, autoimmune diseases and neurodegenerative disorders. The method involves testing the test substances for its ability to inhibit the activity of GADD34-like (GADD34L), also referred to as eIF2alpha-specific regulatory subunit of phosphatase, and identifying the test substance which inhibits the activity of GADD34L. The present sequence is mouse GADD34L cDNA.

XX

SQ Sequence 5468 BP; 1321 A; 1272 C; 1366 G; 1509 T; 0 U; 0 Other;

Query Match 45.0%; Score 1324; DB 12; Length 5468;

Best Local Similarity 72.5%; Pred. No. 0;

Matches 1964; Conservative 0; Mismatches 660; Indels 84; Gaps 16;

Qy 96 CCATCGCCGAGGCGGTGGGGGACTCCTATCCATGGTGTGAAGCGTCGAGCCGACTAGG 155

Db 151 CCGTCGCGCGCGGAGGAGGCTCTTCTATGTGGAGCGATCTCACAGGCGCTAGG 210

Qy 156 GAACCTCTTCCC-CGCCAGGATGGAAGTGCATCAGTCGCGCTATTTCGGCGGGCTGT 214

Db 211 ACGTCTCTCTTCCCTAGCCGGATGGACCTAACCGCGGTCCGACCGCTTTCGGCGGGCTC 270

Qy 215 TCTTCCCTGTGTTCTGCGCGCGCTGCGCATTCGCTGCGCTCTGTGGCTTTTCTGTGG 274

Db 271 TGGGCGCTCGGTGCAGCACTCGTTGGGGAAGCGCGCTCTCTGGGCTCTCTCTGCGG 330

Qy 275 CTCGAAGATCGGCTTGAGCAGCAGCCACCGCTGGGCAAGCGCGAGACTCTGTAGGCT 334

Db 331 CGCGGGAATCGGACTGCAGTACCACTCCGCTGGCTGGCAAGCGCGAGACTGTGTAGACC 390

Qy 335 TCCTCGAATCCGTCGACCTCAGCGCTGAGCGCGCGCGGCTTACCTGAGAGACTGTC 394

Db 391 TCGGATCCAGCTCGCTGACGCGCTGAGCTCTGTCTCTCTCTGTCTGAGAAGCGGCC 450

Qy 395 AAGAAAAAGAGATGGAGCGGCGGACAGCGGATCGCGGAAACGGCTTGGCCCTCGGGCG 454

Db 451 AAG-GAAGAGATGGAGACAGGACGCAAGGCGCGGAGCGGCTTGGCCCTCGGGCTG 509

Qy 455 GGCTTCGGTTCCTGGCCACCCCTTTTTCCTCGGCGATCGCAAGCAGGCTCTTCTAAGTTC 514

Db 510 GGCTTCCTGTTCCGGCTGCCC---TTCCTTCGGGATCGCACGCTCTCTTCGGAGTTC 566

Qy 515 CCGAGCGCTCTTGGCCCGGAAACTCCGGGAACCCACACTGCTTCTCTGCCAGCCCC 574

Db 567 CCGCGCGCTTCTCTCTGACAAAAATCCGGGAA-----CTCCGCTCTGCCCC 611

Qy 575 GAGACTCGGTCAGTTACTGCGAGAAATCTCTCCAGCTCTTGGCGCGCTCCCGGA 634

Db 612 GAGGTCGGAACAGGTAAGTCTGACAAATTTCTTCTCAGCTCTTGGCCCTGCTCCCTAGC 671

Qy 635 TTGCTTCAGAAAGTGTCTAATTTGGAGCCAACTTTTGGTGGAAATGTTTCCGACAGATGG 694

Db 672 CTATTCAGAAAGTGTCTGCTTTGGAGCCAGCTTTTCGGGGGCGCTGATTCTTACAGATGG 731

Qy 695 CTAGATTTTGTGAGTCTATCAGGCGGCTGAGAGCCCTGAGGAGCGGGAACCAAGCC 754

Db 732 CTAGATTTTTCGCGCAAGTTTACAGCGCCCTGAGAGCTTCGAGAGGACGGAGGAACTCTGAC 791  
 Qy 755 GCCCCACAGCCGAGAAATCTTTGAGTT--CGCTGAGCTCGACTCCTC---AGACCCCTC 810  
 Db 792 GCTCCACAGGTGAGAGTCTCTGAGTTTACACTGGCGCTGGACTCTTCGGAAGACTCGC 851  
 Qy 811 GGTCAACAGTCCCTTGGATTGGCTAGAGGAGGGATCCACTGGCAATACTCGCCCCCAGA 870  
 Db 852 GTCGTCAGTACTCTTGCACTTGCTAGAGGAGGACTCCAAGTGGAGTGTCTGCTCTCAGA 911  
 Qy 871 CTTAAATTTGGAGCTTAAGGCCAAGGGAAGTCTTTGGACCTCGAGACAGGCTTTTCT 930  
 Db 912 -CTGGAAGTTAAACTCAAGGCCCCAGGAAGAGCTTTAGACTCTGAGCGCCCACTTTCT 970  
 Qy 931 CTTAGCAGCAGCTGTGGGAGTGGAGCTGTTCGCCAGTAGCTTTCAATCCCGTCTGTA 990  
 Db 971 CTTGAGCAGCAGCTGTGGGAGTGGAGTGTCTGCCAGTAGCTTTCAAGCTGFTTAGT 1030  
 Qy 991 CTTAAACCGGGAACCTTGGCTCTTCGCCCTCTGGGCTCTAAACATTTCAACGCATAGACAA 1050  
 Db 1031 CTCCACCGAGAACTTGACTCTTCATCTCTCGGCTCTGAGCGTTTCAGAGCTTAGGTAA 1090  
 Qy 1051 TTTAGTGTGATCTGATTTGCTGAAACCTTCTCTACCTGACTGCTTTCTTAGGCTAGA 1110  
 Db 1091 TTTCAAGGTAGTTTCTATCTCTGAAACCTTTCTACCTGGACTACTCTTCCCGAGTTAGG 1150  
 Qy 1111 AGTCAGCTATCAGAACAGTGTGGAATAGCGAGGTAGTGGCTTCCAGACACTAACCCC 1170  
 Db 1151 GCTGCGCTGTCAGAGCAGCGCTGGAGGTGGCCAGTTGTGGGTTTCCGAACTAACCCC 1210  
 Qy 1171 AGAGAGCAGCTGCTCGAGAGGAGCAATTTGTCAATCCCGCGCTGAGTCGAGAACTCAT 1230  
 Db 1211 AGAGAGCTGTATCTTTCTGAGATGGTTGTCACTCTCAGCGCTTGGCGGAGAGATGTC 1270  
 Qy 1231 TCGGCGCTGTGGAGGATGTCACCTCTTTCTACGGAAGGCTTACAGAAAATTCACCA 1290  
 Db 1271 GGCACACCGCTTGGAGAAGGTGTCGCTCTCTCTACAGAAAGGCTTGGCGGAAATCCACA 1330  
 Qy 1291 TCTTCGCTGAAACGGCTGGAATTCCTTCAACAGGCTAACAGGGGCAAGATTATACCCAC 1350  
 Db 1331 CTTTGTATGAAACGGCTAGAAATTCCT---CGAGCTAACAAAGGCGAAGTTTACCCAC 1387  
 Qy 1351 CCTGACAGGATATGGCTACACAGCTTGGAGGAGGAACACAGCTTCTCCGATGGA 1410  
 Db 1388 CCTGACCAAGATATAGCTATCATAGCTTGGAGGAGGAACATAACCTTCTCCGATGGA 1447  
 Qy 1411 TCCAAACACTCAGAGATAACCCACACAGTTTGTCTGCTGCTGAGACATTCCTGG 1470  
 Db 1448 CCCACAACATTTGCAAGATAACCCACACAGCGGTGTCCCTGTGCGACAGGCCC--- 1504  
 Qy 1471 AAACACCCAGGAATCCACTGGAAGAAAATAAGAAATTAATTAACACAGAGGTTCCACTTGC 1530  
 Db 1505 -----GGAGCCCACTGAGAAAACCCAGAAATGGTGATTCAAGAAATTTAC--- 1551  
 Qy 1531 TTTGGAAGAAGAGAGCCCTTCTGAGGCTGTCCATCTAGTGAGATPACCTATGGAAAAGGA 1590  
 Db 1552 -----AGAGCCCCAGGGAAGCAGTCTGTTTGTGAATTAACCGTGGAAAAGA 1600  
 Qy 1591 GCCTGAGAGGCGCAATAAGTGTAGTTGATTAATTAATTAACACAGAGTGTGACCTTCCAT 1650  
 Db 1601 ATGTGAAGAGGACCACTAATATGCACTGACCTCTCAGATAGAGGAGAGGCTTTCTGT 1660  
 Qy 1651 TTTCTCCAGACAGCTGTAGTAAACAACTAGATGATTAATTTTGGAGGTGCAATCCAG 1710  
 Db 1661 TTCTACACAGACAGTTTGTAGCAACAACTGATGATTAATTTTGGAGGCGCCCCAG 1720  
 Qy 1711 TGACCTGGAACAAAGTTCTGATPCCAGAAAGGTGAGGATTTGGGATGAGGAAGCTGAGGATGA 1770  
 Db 1721 TGACTTGAAGCCAGCTCTGATTTCTGAAAGTGAAGTTGGGCGGAGGAACTGAGGACGA 1780  
 Qy 1771 TGGTTTTGATAGTGTAGTCTGCTGAGACTCAGACTTCAAGACCTTCAAGACCTTGAAGGCT 1830  
 Db 1781 TGGCTTTTGTATAGCGATGGCTCCCTGTCTGAAATCAGACGTGGAACAGGACTCGGAAGGCT 1840

Qy 1831 TCACCTTTGGAACCTCTTTCTGAGTGTAGATCCCTTATAATCCCCAGAACTTTTACAGCAAC 1890  
 Db 1841 TCACCTTTGGAACCTCTTTCCACAGTGTAGATCCCTTACAAACCCCAAAACTTTTACAGCCAC 1900  
 Qy 1891 AATTACAGACTGCTGCCAGAAATTTGTTCTGGAAGACCTTCTGATTTACAGAAAGGATTTGTC 1950  
 Db 1901 GATTACAGCGCTGCCAGAAATTTGCCCCAGAGAGCCATCAGATTCAGGAGCATCTCTGCTC 1960  
 Qy 1951 TGGCAAGCTCTGATCTAGAGAAATTCCTCCAGTCTGGAAGCCTTCTGAGACCCCTGAGCA 2010  
 Db 1961 TGGCAGCTGTGGTGTAGGGA---GCTGTCAAGAGGAGCCCTTCCGAGAGCCCCGACCA 2017  
 Qy 2011 TAGTTCTGGGAGGAAAGATGACTGGGAATCTAGTGCAGATGAAGCAGAGAGTCTCAAACT 2070  
 Db 2018 TAGTTCCGGGAGGAAAGATGACTGGGAACCGAGTGCAGATGAAGCAGAGAACTCTTAAAT 2077  
 Qy 2071 GTGAACTCATTTCTGTAATTTCTGATGACCCCTCAACCCCTTTAAATTTAAAGGCTCCTTT 2130  
 Db 2078 GTGAACTCTTTCTGTCATTTCTGAGGACCCCTCAACCTTTTAAATTTTAAAGGCTCCTTT 2137  
 Qy 2131 TCAAACATCAGGGGAAATGAGAAAGGCTGTGCTGACTCAAGACCCCATCTGAGTCCAT 2190  
 Db 2138 TCAACCGTCAAGGGAAGAAATTTGGAAGGCGCTCAGGACTCAAGAGGCTCTTCTGAGGTAC 2197  
 Qy 2191 TGTGCCCATTTCTGAGTGTCACTTACTTCTTCTGTAAGGTGCAGCTGTTGGGGAGCCA 2250  
 Db 2198 AGTGGCTTCTCTGGCCATCATACCTTACTTCTTGTGAAGGCCAGCTGTTAGAGAGCCA 2257  
 Qy 2251 AGAAAGTGAATGTCCAGACTCGGTAACGCTGAGTGTCTTTCTGGAAGGAGACACACACA 2310  
 Db 2258 AGAAGATAATTTGTCAGGCTGTGGCTGTGGGTGAGGCTCTTGTGGAAGAAATACACCCA 2317  
 Qy 2311 TGTCAAGGAAAGAAAGGTAACTTCTTGAAGAGTTACTGAGTATATATAGTGTGGA 2370  
 Db 2318 TATCAAGAGAAAAGGTAACTTCTTGAAGAGTTACTGAGTATATATAGTGTGGA 2377  
 Qy 2371 TGAGGATCGCAAGAGCAATGGGAAGAAATTTGCAAGGAGTGGATGCAAGTTCCAGAAACG 2430  
 Db 2378 TGAGGATCGCAAGAGCAATGGGAAGAAATTTGCAAGGAGTGGATGCAAGTTCCAGAAACG 2437  
 Qy 2431 AATTCAAGAAACAGAAAGATGCTATTTGGATATTTGCTGACATTTTGAACACAGAGAAAGAT 2490  
 Db 2438 AATTCAAGAAACAGAAAGTTGCCATTTGGCTACTGCTTGGCCCTTTGAGCACAGAGAAAAAT 2497  
 Qy 2491 GTTTAATAGACTCCAGGGAACATGCTTCAAGGACTTAAATGTTCTCAAGCAATGTTGAGT 2550  
 Db 2498 GTTTAATAGACT---GAGGATCGAGTCAAGGAGCTTACTGTTGTACAGCAATGTTAAGA 2553  
 Qy 2551 TGGCAGCTCTGTAGTCTTAGCTAGCATACACTACCTCTTACTGAGAGGTGCTTTTAAAA 2610  
 Db 2554 AGTGA-ACAGCTGCAACCCGTGCCACTCTGTCTCTTACTTGAGA-GTTTTCCCTTAAAA 2611  
 Qy 2611 ACAAACTTGGCAGCTGTCTTTGACATTTTGTGATTTTGTAGAGGAAATGTAATTTGATCT 2670  
 Db 2612 ACAAACTTGGCAGCTGTCTTGGACATGTTTTTAAAGAAACCACTTTGTATCTAGAGATG 2671  
 Qy 2671 AGTTTAATTTTGTGCAACATATCCCACTCAGAAACATTCAGGTGTTTGAAGCCAGCC 2730  
 Db 2672 CAGTTTGAATTTTGGGTAAATGTTCTCATTAGAA-----ACACCAACT 2717  
 Qy 2731 CTGATAATGAAGGATGAAGTGTGATTTCTAATCTCCCTTTTGTGATTTAGTTGGAT 2790  
 Db 2718 CCGATAATGAAGAAAT-----CTCTTATCTGTAATCTCTCTTTTCCTATTTAGTTGGAT 2771  
 Qy 2791 GTGCTTTT 2798  
 Db 2772 GTGGGTTT 2779



KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.  
XX Homo sapiens.  
XX EP1074617-A2.  
XX 07-FEB-2001.  
XX 28-JUL-2000; 2000EP-00116126.  
XX 29-JUL-1999; 99JP-00248036.  
XX 27-AUG-1999; 99JP-00300253.  
XX 11-JAN-2000; 2000JP-00118776.  
XX 02-MAY-2000; 2000JP-00183767.  
XX 09-JUN-2000; 2000JP-00241899.  
XX (HELI-) HELIX RES INST.  
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX WPI; 2001-318749/34.  
XX  
XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.  
XX  
XX Claim 1; SEQ ID NO 3458; 2537pp + Sequence Listing; English.  
XX  
XX The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention  
XX  
XX Sequence 772 BP; 134 A; 256 C; 217 G; 161 T; 0 U; 4 Other;  
SQ  
Query Match 25.1%; Score 738; DB 4; Length 772;  
Best Local Similarity 98.8%; Pred. No. 5.7e-209; Mismatches 8; Indels 1; Gaps 1;  
Matches 752; Conservative 0;  
QY 1 ATTTTGGCTTCGTTCCACCGCAGCCGGCTTACCGAGTCCTTCGGGTATCGCGTTG 60  
DB 1 ATTTTGGCTTCGTTCCACCGCAGCCGGCTTACCGAGTCCTTCGGGTATCGCGTTG 60  
QY 61 CTCAGGGGCTTTCAACCTCTCTAGTCGGAAACCATCGCCGAGCGCGTGGGGGACT 120  
DB 61 CTCAGGGGCTTTTCAACCTCTCTAGTCGGAAACCATCGCCGAGCGCGTGGGGGACT 120  
QY 121 CCTATCCATGGTGTGAAGCGTCGAGCGGACTAGGGAACCTCTTCCCGCCGAGGATGGA 180  
DB 121 CCTATCCATGGTGTGAAGCGTCGAGCGGACTAGGGAACCTCTTCCCGCCGAGGATGGA 180  
QY 181 AGTCGCATCAGTCGCGCGCTATTGCGGGGCTGTTCTTCCCTGTGTTCTGCGCGCGCTG 240

DB 181 AGTCGCATCAGTCGCGCGCTATTGCGGGGCTGTTCTTCCCTGTGTTCTGCGCGCGCTG 240  
QY 241 CCGCATTCGCTGCCCTCTGTGGCTTTTCTGCTGGCTCGAGATCGGCTGAGCAGCGAC 300  
DB 241 CCGCATTCGCTGCCCTCTGTGGCTTTTCTGCTGGCTCGAGATCGGCTGAGCAGCGAC 300  
QY 301 GCCACCGCTGGGCAAGGCCGAGACTCTGTAGGCTTCTCCGAATCCCGTCCAGCTCCAGC 360  
DB 301 GCCACCGCTGGGCAAGGCCGAGACTCTGTAGGCTTCTCCGAATCCCGTCCAGCTCCAGC 360  
QY 361 CGCTGAGCGCGCGGCTTACCTGAGAGACTGTCAAGAAAAGAGATGGAGCGGGGAC 420  
DB 361 CGCTGAGCGCGCGGCTTACCTGAGAGACTGTCAAGAAAAGAGATGGAGCGGGGAC 420  
QY 421 AGCGGATCGCGGAAACGGCTTGGCCCTCGGGGGGCTTCGGTTCGGCCACCTTTT 480  
DB 421 AGCGGATCGCGGAAACGGCTTGGCCCTCGGGGGGCTTCGGTTCGGCCACCTTTT 480  
QY 481 CCCTCGCGGATCGCAAGCAGGCTCTTAAAGTTCCCGACGCTCTTGCGCCGGAAAATC 540  
DB 481 CCCTCGCGGATCGCAAGCAGGCTCTTAAAGTTCCCGACGCTCTTGCGCCGGAAAATC 540  
QY 541 CGGAAACCCACACTGTTCTGCTGCGCGGAGCTCGGGTCAAGTCTGAGCGAA 600  
DB 541 CGGAAACCCACACTGTTCTGCTGCGCGGAGCTCGGGTCAAGTCTGAGCGAA 600  
QY 601 ACTGCTCTCCAGCTCTTGGCGGCTCCCGGATTCCTCAGAAAGTGTAAATTTGAG 660  
DB 601 ACTGCTCTCCAGCTCTTGGCGGCTCCCGGATTCCTCAGAAAGTGTAAATTTGAG 660  
QY 661 CCAACTTTTCGTTGGAATGTTTCCGACAGATGGCTAGATTTTCTGAGTCTAC-AGCG 719  
DB 661 CCAACTTTTCGTTGGAATGTTTCCGACAGATGGCTAGATTTTCTGAGTCTACAGCG 720  
QY 720 CCCTGAGAGCTTGAAGGAGCGGAGAAACAGCGCGCGCC 760  
DB 721 CCCTGAGAGCTTGAAGGAGCGGAGAAACAGCGCGCGCC 761  
RESULT 7  
AAH11666/c  
ID AAH11666 standard; cDNA; 543 BP.  
XX  
XX AC AAH11666;  
XX 26-JUN-2001 (first entry)  
XX  
XX Human cDNA clone (3'-primer) SEQ ID NO:8501.  
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.  
XX Homo sapiens.  
XX EP1074617-A2.  
XX 07-FEB-2001.  
XX 28-JUL-2000; 2000EP-00116126.  
XX 29-JUL-1999; 99JP-00248036.  
XX 27-AUG-1999; 99JP-00300253.  
XX 11-JAN-2000; 2000JP-00118776.  
XX 02-MAY-2000; 2000JP-00183767.  
XX 09-JUN-2000; 2000JP-00241899.  
XX (HELI-) HELIX RES INST.  
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX WPI; 2001-318749/34.  
XX  
XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-





QY 1679 CTGATAGATTATATTTTGGGAGGTGCATCCAGTGACCT 1716



Db 517 GCGTTAAAGCCCATTTTCATTTCTAAATTAATGTCGCTATCTG 469

RESULT 11  
ABX52167 standard; cDNA; 399 BP.

XX AC ABX52167;  
XX DT 25-FEB-2003 (first entry)

XX DE Bovine EST associated with lactation/muscle/fat deposition #2096.

XX KW Bovine; ss; EST; expressed sequence tag; lactation; LMFD;  
KW muscle deposition; fat deposition; genome mapping; gene identification;  
KW gene analysis; cattle breeding.

XX OS Bos Taurus.

XX PN US2002137160-A1.

XX PD 26-SEP-2002.

XX PF 26-OCT-2001; 2001US-00983965.

XX PR 17-DEC-1998; 98US-0113678P.

XX PR 15-DEC-1999; 99US-00465231.

XX PA (BYAT//) BYATT J C.

XX PA (MATH//) MATHIALAGAN N.

XX PA (TAON//) TAO N.

XX PA (WARR//) WARREN W C.

XX PI Byatt JC, Mathialagan N, Tao N, Warren WC;

XX WPI; 2003-102386/09.

XX PT Purified nucleic acid molecules, useful for genome mapping, gene  
PT identification and analysis, cattle breeding or preparation of constructs  
PT for cattle gene expression and genetically improved cattle.

XX PS Claim 2; SEQ ID NO 2096; 38pp; English.

XX CC The invention relates to a purified nucleic acid molecule associated with  
CC lactation or muscle and fat deposition (designated LMFD), derived from  
CC cattle, and the LMFD nucleic acid can specifically hybridize to a second  
CC nucleic acid molecule comprising any of 5912 nucleotide sequences,  
CC appearing as ABX50072-ABX55983, or complements of them. Also included are  
CC ; (1) a transformed cell having a nucleic acid comprising an LMFD nucleic  
CC acid linked to a promoter and a 3' non-translated sequence that  
CC functions in the cell to cause termination of transcription and addition  
CC of polyadenylated ribonucleotides to a 3' end of the mRNA molecule; and  
CC (2) determining a level or pattern of a molecule in a bovine cell or  
CC tissue comprising: (a) incubating a marker nucleic acid (comprising any  
CC of the 5912 nucleic acid sequences or its complement or fragment) with a  
CC complementary nucleic acid molecule obtained from the bovine cell or  
CC tissue, where hybridisation between the marker nucleic acid and the  
CC complementary nucleic acid permits the detection of the molecule; and (b)  
CC detecting the level or pattern of the complementary nucleic acid, where  
CC the detection of the complementary nucleic acid is predictive of the  
CC level or pattern of the molecule. The LMFD nucleic acid is used for  
CC determining a level or pattern of a molecule in a bovine cell or tissue.  
CC It is useful for genome mapping, gene identification and analysis, cattle  
CC breeding, preparation of constructs for use in cattle gene expression, or  
CC for genetically improving cattle. The present sequence is one of the 5912  
CC bovine LMFD EST (expressed sequence tag) nucleic acids. Note: The present  
CC sequence was not shown in the specification but was obtained in  
CC electronic format from the USPTO web site:  
CC seqdata.uspto.gov/sequence.html?docID=20020137160

XX Sequence 399 BP; 133 A; 62 C; 82 G; 122 T; 0 U; 0 Other;

Query Match 6.8%; Score 198.6; DB 8; Length 399;  
Best Local Similarity 83.8%; Pred. No. 4.8e-48;  
Matches 249; Conservative 0; Mismatches 44; Indels 4; Gaps 2;  
  
QY 2323 AAAGGTAACCTTCCTTGAAGAAGTTACTGAGTATTATATAGTGGTGTGATGAGGATCGCAA 2382  
DB |||||||  
QY 102 ATAGGTAACCTTCCTTGAAGAAGTTACTGAGTATTATATAGTGGTGTGATGAGGATCGAAA 161  
DB |||||||  
QY 2383 AGGACCATGGGAAGAAATTTGCAAGGGATGATCGAGTTCACAGAAACGAATTCAGAAAC 2442  
DB |||||||  
QY 162 AGGACCATGGGAAGAAATTTGCAAGGGATGATCGAGTTCACAGAAACGAATTCAGAAAC 221  
DB |||||||  
QY 2443 AGAAGATGCTATTGGATATTGCTTGACATTTGACACAGAGAAAGATGTTTATAGA-- 2500  
DB |||||||  
QY 222 AGAAGAAGCTATTGGATACCTGCTTGACATTTGAGCAGAGAGAAAGATGTTTTAAATAAA 281  
DB |||||||  
QY 2501 -CTCCAGGGGAACATGCTTCAAGGAGCTTAAATGTTCTCAAGCAATGTTGAGTTCG- 2558  
DB |||||||  
QY 282 CTTCAAGCAACATGTTTCAAGGATCTTAATGTTTTCGAGCATGTTTATAGATGATTCGAC 341  
DB |||||||  
QY 2559 TGTAGTCTTAGCTAGCATACATCTTACCTCTGAGAGGTGCTTTTAAAAACAA 2615  
DB |||||||  
QY 342 AGCCTCTGGCCCTAGCATACATCTTACCTCTGAGAGGTACTTTTAAAAACCAA 398  
DB |||||||

RESULT 12

AAC06938  
ID AAC06938 standard; cDNA; 195 BP.

AC AAC06938;

XX DT 06-OCT-2000 (first entry)

XX DE Human secreted protein 5' EST, SEQ ID NO: 11013.

XX KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
KW Gene therapy; chromosome mapping; ss.

XX OS Homo sapiens.

XX PN EP1033401-A2.

XX PD 06-SEP-2000.

XX PF 21-FEB-2000; 2000EP-00200610.

XX PR 26-FEB-1999; 99US-0122487P.

XX PA (GEST ) GENSET.

XX PI Dumas Milne Edwards J, Duclert A, Giordano J;

XX WPI; 2000-500381/45.

XX CC New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
CC obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for  
CC diagnostic, forensic, gene therapy and chromosome mapping procedures.

XX PS Claim 1; SEQ ID NO 11013; 71pp + Sequence Listing; English.

XX CC The present sequence is one of a large number of 5' ESTs derived from  
CC mRNAs encoding secreted proteins. No ORF has yet been conclusively  
CC identified within the present sequence. The 5' ESTs were prepared from  
CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST  
CC sequences usually correspond mainly to the 3' untranslated region (UTR)  
CC of the mRNA because they are often obtained from oligo-dT primed cDNA  
CC libraries. Such ESTs are not well suited for isolating cDNA sequences  
CC derived from the 5' ends of mRNAs and even in those cases where longer  
CC cDNA sequences have been obtained, the full 5' UTR is rarely included. 5'  
CC ESTs are derived from mRNAs with intact 5' ends and can therefore be used  
CC to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used in  
CC diagnostic, forensic, gene therapy and chromosome mapping procedures.  
CC They are used to obtain upstream regulatory sequences and to design



identity to (a); or degenerate variants of (a). Also included are an isolated polypeptide (b) comprising: sequences encoded by the new polynucleotide; any of the 4 amino acid sequences fully defined in the specification; or sequences having at least 70 or 90% identity to the sequence in (a) or (b)), an expression vector comprising the above polynucleotide operably linked to an expression control sequence, a host cell transformed or transfected with the above expression vector, an isolated antibody, or its antigen-binding fragment, that specifically binds to the above polypeptide, an oligonucleotide that hybridises to the above-mentioned nucleotide sequences under highly stringent conditions, a fusion protein comprising at least one polypeptide cited above, detecting the presence of a cancer in a patient (comprising: obtaining a biological sample from the patient; contacting the biological sample with a binding agent that binds to the polypeptide, or with the oligonucleotide cited above; detecting in the sample an amount of the polypeptide that binds to the binding agent, or an amount of a polynucleotide that hybridises to the oligonucleotide; and comparing the amount of polypeptide, or polynucleotide that hybridises to the oligonucleotide, to a predetermined cut-off value and then determining the presence of a cancer in the patient), a method for stimulating and/or expanding T-cells specific for a tumour protein (comprising contacting T-cells with the above polypeptide, polynucleotide or antigen-presenting cells that express the polynucleotide, under conditions and for a time sufficient to permit the stimulation and/or expansion of T-cells), an isolated T-cell population comprising T-cells prepared by the method, a composition comprising a first component selected from physiological carriers and immunostimulants, and a second component selected from the above polypeptide, polynucleotide, antibody, fusion protein, T-cell population and antigen-presenting cells that express the above polypeptide, stimulating an immune response in a patient (comprising administering to the patient the above composition) treating lung cancer in a patient (comprising administering to the patient the above composition and a diagnostic kit (comprising: at least one oligonucleotide cited above; or at least one antibody cited above and a detection reagent, where the detection reagent comprises a reporter group). The composition and methods are useful in diagnosing, preventing and treating cancer, particularly lung cancer. The present sequence is a lung cancer-associated antigen cDNA of the invention.

Sequence 173 BP; 73 A; 30 C; 28 G; 42 T; 0 U; 0 Other;

Query Match 5.5%; Score 162; DB 10; Length 173;  
Best Local Similarity 99.4%; Pred. No. 2.4e-37;  
Matches 173; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

2625 CTGTCCTTGCACATTTTCTTTTATAGAGGAATGTAACCTGGATCTAGTTTATTTT 2684  
173 CTGTCCTTGCACATTTTCTTTTATAGAGGAATGTAACCTGGATCTAGTTTATTTT 115  
2685 TTTTGCAACATATCCCACTCAGAAACATTGAGGTTTGAAGCCAGCCCTGATATGAAGGA 2744  
114 TTTTGCAACATATCCCACTCAGAAACATTGAGGTTTGAAGCCAGCCCTGATATGAAGGA 55  
2745 TGAACATAGTGATTTCTAATCTCCCTTTTGTGATTTAGTTGATGCTTTT 2798  
54 TGAACATAGTGATTTCTAATCTCCCTTTTGTGATTTAGTTGATGCTTTT 1

RESULT 15  
ADA71938  
ID ADA71938 standard; DNA; 2000 BP.  
XX AC ADA71938;  
XX AC ADA71938;  
XX DT 20-NOV-2003 (first entry)  
XX DE Rice gene, SEQ ID 5263.  
XX KW Plant; bacterial infection; fungal infection; viral infection; rice;  
XX KW gene; ds.  
XX OS Oryza sativa.  
XX

PN WO2003000898-A1.  
XX 03-JAN-2003.  
XX 22-JUN-2001; 2001WO-IB001105.  
XX 22-JUN-2001; 2001WO-IB001105.  
PR (SYGN ) SYNGENTA PARTICIPATIONS AG.  
XX Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;  
PI Katagiri F, Qian S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;  
XX WPI; 2003-175290/17.  
XX Identifying at least one gene involved in plant resistance or response to  
PT pathogenic infection for conferring resistance or tolerance to a plant to  
PT bacterial, fungal or viral infection by determining or detecting plant  
PT gene expression.  
XX Claim 27; SEQ ID NO 5263; 899pp; English.  
XX The present invention relates to a method (M1) for identifying genes  
CC involved in plant resistance or response to pathogenic infection. M1  
CC comprises identifying a gene whose expression is significantly altered in  
CC the incompatible interaction of plant gene expression relative to  
CC expression of the gene in an uninfected plant, in a mutant plant that  
CC does not express a gene associated with response to pathogenic infection,  
CC or in a corresponding incompatible or compatible interaction. (M1) is  
CC useful for conferring resistance to resistance or tolerance to a plant to  
CC bacterial, fungal or viral infection. The present sequence was used to  
CC illustrate the invention.  
XX Sequence 2000 BP; 336 A; 265 C; 284 G; 363 T; 0 U; 752 Other;  
SQ Query Match 2.1%; Score 61.8; DB 8; Length 2000;  
Best Local Similarity 11.7%; Pred. No. 1e-06;  
Matches 110; Conservative 402; Mismatches 414; Indels 11; Gaps 3;  
QY 1705 ATCCAGTACCTGGAAACAGTTCTGATCCAGAGGTGAGGATGGATGAGGAGCTGA 1764  
DB 158 RKXSGSGWGTCTRRGARGSGWSSGAKYKSGSMKMMSSCGRRSSYSRYTGT 217  
QY 1765 GATGATGTTTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1824  
DB 218 SRXYGYKXWYYSASRCMEAYMTTSYACSSYVTCRSKRSMWMMKMRMSRYG 277  
QY 1825 AGGGCTTACCTTTGGAACTCTTTCTGAGTGTAGATCTTATATCCCAAGACTTTAC 1884  
DB 278 YSWSYKMMWCTAYKK--SYSRWCYMYRGGWGRGATRYMGYGRSMAMMYKMYWYRG 334  
QY 1885 AGCAACAATTCAGACTGCTGCCAGAAATTTCTCTGAAGAGCCCTTCGATTCAGAGGA 1944  
DB 335 YKMGKMGWAGRMMSMRMSKRWSS--KACYMYRWRWRMTTTRRRWAKKSRSTRRKKRWKMR 393  
QY 1945 TTTGTCTGCAAGTCTGATCTAGAGAAATTCCTCCAGTCTGGAAGCCCTTCCTGAGACCC 2004  
DB 394 KRXYKMRGYSRWBSCKBARMMKCRSGRAWKMGCGCHTCRMKSYGMWKSWMKWSK 453  
QY 2005 TGAGCATAGTTCTGGGAGGAAGATGACTGTTGGAAATCTAGTGCAGATGAAGCAGAGTCT 2064  
DB 454 YKMSRMYRWRKCKSRTTMMGKTGMMGTMCRCYKRSKMRKCRRRRWRGMYRMRW 513  
QY 2065 CAAACTGTGGAACATCTCTGTAATTTCTGATGACCCCTCAACCCCTTTTAATTTAAGGC 2124  
DB 514 KRYMSARYMTMYCARKKYSYSARARCMYRKGYYWAGMMWKKRYKMTMYKMMWYKR 573  
QY 2125 TCCTTTTAAACATCAGGGGAAATGAGAAAGCTGTCTGATCTCAAGACCCCATCTGA 2184  
DB 574 KYSKSWYKMSYASCMKSKARAGAKMCKRSKWSMSRSMRSKCRKCKASRKSARKY 633  
QY 2185 GTCCATTTGGCCCATTTCTGAGTGTGACACTTACTTTCTTGTAAAGTGAGCTGTTGGG 2244

Db	634	AMMGMTSGSRMSRWKSYTC-----YWRKWSGMSKSTCTWMYMSKYTYAKTGSYWRYYR	686
Qy	2245	GAGCGAAGAAAGTGAATGTCAGAGACTCGGTACAGCGTGACGTTCTTTCTCGGAGGAAGACA	2304
Db	687	YRAMCYMRWYYRYRYSYMTYMAWYTSSTRMAWYTGMYKSGRYWTSWYKYCKCSWKYRSM	746
Qy	2305	CACACATGTCMAAAGAAAAAGGTAACTTCCTTTGGAAGAAGTTACTGAGATTATATAAG	2364
Db	747	WYYSWMAWKTWQWRRYATRMHMMWYRYSNMKYTWCTWMGYWYWWRTYTMQRYYMYKC	806
Qy	2365	TGGTGATGAGGATCGCAAGGACCATCGGAAGAATTTGCAAGGATCGATGCAGGTTTCCA	2424
Db	807	TKTYWYHSATYWTGTWAAWMAWKTMRMGWGTGATRGSRARKARYWYWKXATWCAIKRWMTX	866
Qy	2425	GAACAGAAATTCAGAAAACAGAAGATGCTATTGGATATTGCTTGCAATTTGAAACACAGAGA	2484
Db	867	GKGAKWATWMAKAWRYKYWSMRWYKYKTRTRRYKTCWKRWGSWAYWRMWWKGS	926
Qy	2485	AAGNAATGTTTAAATAGACTCCAGGGAACATGCTTCAAAGGACTTAATGTTCTCAAGCAATG	2544
Db	927	KQMMWMMGGRWGWTXYWYCTTWKMAAGRATKYMCCAGWAMYSWYRTYWMRTWRWM	986
Qy	2545	TTGAGTTGGCAGCCCTGTAGTCTTAGCTAGCATACATCACTCTTTACCTGAGAGGTTGCTTT	2604
Db	987	WASSRTAKRWARMWKTWMSKSYARAYWMAWGCACCTACACACTACACAGAAATGTTGT	1046
Qy	2605	TTAAAAACAAATCTGGGAGCTGTCTTTTGACATTTT	2641
Db	1047	TTTATTTTCAATTCATTTGGTACTCTCTATGTAAGTGT	1083

Search completed: September 15, 2005, 16:13:36  
Job time : 1436.17 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 15, 2005, 12:27:30 ; Search time 439.026 Seconds  
(without alignments)  
10965.010 Million cell updates/sec

Title: US-10-650-482-1

Perfect score: 2942

Sequence: 1 attttggcttcctccac.....aaattatgtgcttatctg 2942

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

Issued Patents NA:\*\*

1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq.\*

2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq.\*

3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq.\*

4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq.\*

5: /cgn2\_6/ptodata/1/ina/PTCUS\_COMB.seq.\*

6: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	219	7.4	226	4	US-09-621-976-8778
2	218	7.4	218	4	US-09-513-999C-33878
3	195	6.6	195	4	US-09-513-999C-11013
4	51	1.7	7218	1	US-08-232-463-14
5	45.4	1.5	1141	4	US-09-806-708B-22
6	42.6	1.4	45365	4	US-09-949-016-11893
7	42.6	1.4	45456	4	US-09-949-016-17007
8	42	1.4	364	4	US-09-621-976-17202
9	41.4	1.4	399	4	US-09-806-708B-22
10	40.8	1.4	1141	4	US-09-949-016-15525
11	40.4	1.4	117807	4	US-09-949-016-13865
12	40.2	1.4	1468	4	US-09-949-016-3573
13	40.2	1.4	1560	4	US-09-023-855-1177
14	40.2	1.4	4252	4	US-09-919-039-334
15	39.6	1.3	7218	1	US-08-232-463-14
16	39.4	1.3	390416	4	US-09-949-016-16923
17	39.2	1.3	12526	4	US-09-949-016-13865
18	39.2	1.3	23439	4	US-08-956-171E-38
19	39.2	1.3	23439	4	US-08-781-986A-38
20	39.2	1.3	161607	4	US-09-949-016-12210
21	38.8	1.3	112705	4	US-09-949-016-15630
22	38.6	1.3	119979	4	US-09-949-016-12309
23	38.6	1.3	19980	4	US-09-949-016-13533
24	38.2	1.3	474	4	US-09-621-976-18033
25	38.2	1.3	90876	4	US-09-949-016-13271
26	38	1.3	7409	4	US-09-949-016-12689
27	38	1.3	7409	4	US-09-949-016-15978
28	37.8	1.3	34261	4	US-09-949-016-16908
29	37.8	1.3	49309	4	US-09-949-016-12680
30	37.8	1.3	49309	4	US-09-949-016-13084
31	37.8	1.3	85368	4	US-09-949-016-12211
32	37.8	1.3	174639	4	US-09-949-016-16509
33	37.8	1.3	462589	4	US-09-949-016-12900
34	37.8	1.3	476044	4	US-09-949-016-12412
35	37.8	1.3	199	4	US-09-513-999C-11028
36	37.6	1.3	410	4	US-09-513-999C-1224
37	37.6	1.3	410	4	US-09-471-276-750
38	37.6	1.3	410	4	US-09-107-532A-1647
39	37.6	1.3	656	4	US-09-949-016-36413
40	37.4	1.3	601	4	US-09-949-016-46425
41	37.4	1.3	601	4	US-08-446-137B-1
42	37.2	1.3	3279	3	US-09-949-016-15545
43	37	1.3	74881	4	US-09-949-016-12286
44	37	1.3	74914	4	US-09-949-016-14711
45	37	1.3	150780	4	US-09-949-016-13379
46	37	1.3	173952	4	US-09-328-352-1597
47	36.8	1.3	3267	4	US-09-270-767-3435
48	36.6	1.2	543	4	US-09-614-221A-342
49	36.6	1.2	543	4	US-09-949-016-16703
50	36.6	1.2	2295	4	US-09-949-016-15999
51	36.6	1.2	12323	4	US-09-949-016-15000
52	36.6	1.2	46288	4	US-09-949-016-13747
53	36.6	1.2	46311	4	US-09-248-796A-7350
54	36.6	1.2	26465	4	US-09-621-976-16556
55	36.4	1.2	402	4	US-09-949-016-17028
56	36.4	1.2	430	4	US-08-906-156A-23
57	36.4	1.2	88557	4	US-08-906-156A-41
58	36	1.2	414	3	US-09-746-801A-36
59	36	1.2	429	3	US-09-134-000C-3119
60	36	1.2	510	4	US-08-906-156A-68
61	36	1.2	576	4	US-09-140-749-27
62	36	1.2	586	3	US-09-902-540-161
63	36	1.2	587	3	US-08-791-115B-22
64	36	1.2	970	3	US-09-358-381-1
65	36	1.2	970	4	US-09-577-902-1
66	36	1.2	1161	4	US-09-140-749-64
67	36	1.2	1194	4	US-07-991-867B-41
68	36	1.2	1209	4	US-08-544-332-41
69	36	1.2	1212	3	US-09-370-861A-41
70	36	1.2	1212	3	US-08-906-156A-11
71	36	1.2	1212	4	US-08-791-115B-2
72	36	1.2	1689	2	US-09-140-749-64
73	36	1.2	1689	3	US-07-991-867B-1
74	36	1.2	1689	3	US-08-544-332-41
75	36	1.2	1773	3	US-09-370-861A-41
76	36	1.2	3160	3	US-08-906-156A-11
77	36	1.2	3160	4	US-09-140-749-27
78	36	1.2	8457	1	US-07-991-867B-1
79	36	1.2	8457	2	US-08-544-332-1
80	36	1.2	8457	3	US-09-370-861A-1
81	36	1.2	47375	4	US-09-949-016-15420
82	36	1.2	122772	4	US-09-949-016-14132
83	35.8	1.2	474	4	US-09-621-976-18033
84	35.8	1.2	832	4	US-09-621-976-2813
85	35.8	1.2	6664	4	US-09-949-016-14970
86	35.8	1.2	19932	2	US-08-477-451-25
87	35.6	1.2	601	4	US-09-949-016-68090
88	35.6	1.2	832	4	US-09-621-976-2813
89	35.6	1.2	580073	4	US-08-545-528B-1
90	35.4	1.2	601	4	US-09-949-016-74411
91	35.4	1.2	601	4	US-09-949-016-74412
92	35.4	1.2	601	4	US-09-949-016-204315
93	35.4	1.2	2311	4	US-09-774-528-35
94	35.4	1.2	120609	4	US-09-949-016-13915
95	35.4	1.2	268449	4	US-09-949-016-17244
96	35.2	1.2	756	4	US-09-248-796A-4637
97	35.2	1.2	3084	4	US-09-762-724-7
98	35.2	1.2	3090	4	US-09-762-724-5
99	35.2	1.2	118067	4	US-09-497-855A-32
100	35	1.2	409	4	US-09-621-976-1242

Sequence 16908, A  
Sequence 14611, A  
Sequence 12680, A  
Sequence 13084, A  
Sequence 12211, A  
Sequence 16509, A  
Sequence 12900, A  
Sequence 12412, A  
Sequence 11028, A  
Sequence 1224, A  
Sequence 1224, A  
Sequence 750, App  
Sequence 1647, App  
Sequence 16413, A  
Sequence 46425, A  
Sequence 1, Appl1  
Sequence 15545, A  
Sequence 12285, A  
Sequence 14711, A  
Sequence 13379, A  
Sequence 1597, App  
Sequence 3435, App  
Sequence 18717, A  
Sequence 342, App  
Sequence 16703, A  
Sequence 14999, A  
Sequence 15000, A  
Sequence 13747, A  
Sequence 7350, App  
Sequence 16656, A  
Sequence 17028, A  
Sequence 23, Appl  
Sequence 41, Appl  
Sequence 36, Appl  
Sequence 3119, App  
Sequence 20, Appl  
Sequence 68, Appl  
Sequence 16, Appl  
Sequence 5888, App  
Sequence 161, App  
Sequence 22, Appl  
Sequence 1, Appl1  
Sequence 1, Appl1  
Sequence 64, Appl  
Sequence 41, Appl  
Sequence 41, Appl  
Sequence 11, Appl  
Sequence 2, Appl1  
Sequence 1, Appl1  
Sequence 1, Appl1  
Sequence 1, Appl1  
Sequence 1, Appl1  
Sequence 15420, A  
Sequence 14132, A  
Sequence 18033, A  
Sequence 2813, App  
Sequence 14970, A  
Sequence 25, Appl  
Sequence 68090, A  
Sequence 2813, App  
Sequence 1, Appl1  
Sequence 74411, A  
Sequence 74412, A  
Sequence 204315, A  
Sequence 35, Appl  
Sequence 13915, A  
Sequence 17244, A  
Sequence 4637, App  
Sequence 7, Appl1  
Sequence 5, Appl1  
Sequence 1242, App

## ALIGNMENTS

## RESULT 1

US-09-621-976-8778  
; Sequence 8778, Application US/09621976  
; Patent No. 6639063  
; GENERAL INFORMATION:  
; APPLICANT: Dumas Milne Edwards, J.B.  
; APPLICANT: Giordano, J.Y.  
; APPLICANT: Duclert, A.  
; TITLE OF INVENTION: ESPs and Encoded Human Proteins.  
; FILE REFERENCE: GENSET.054PR2  
; CURRENT APPLICATION NUMBER: US/09/621,976  
; CURRENT FILING DATE: 2000-07-21  
; NUMBER OF SEQ ID NOS: 19335  
; SOFTWARE: Patent.pm  
; SEQ ID NO 8778  
; LENGTH: 226  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: 68..76..78  
; OTHER INFORMATION: n=a, g, c or t  
US-09-621-976-8778

Query Match 7.4%; Score 219; DB 4; Length 226;  
Best Local Similarity 96.4%; Pred. No. 2.3e-56;  
Matches 217; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 2 TTTTGGGCTTGGCTTCCACCGGACAGCGGGCTACCCAGTCTTCCGGTATCGGTTGC 61  
DB 2 TTTTGGGCTTGGCTTCCACCGGACAGCGGGCTTCCGGTATCGGTTGC 61  
QY 62 TCAGGGGCTTTTCAACCTCTGTAGTCGGAACCATCGCGAGCGCGTGGGGGACTC 121  
DB 62 TCAGGGGCTTTTCAACCTCTGTAGTCGGAACCATCGCGAGCGCGTGGGGGACTC 121  
QY 122 CTATCCATGGTGTGAAGCGTCGAGCCGACTAGGGAACCTCTTCCCGCCAGGATGGA 181  
DB 122 VMACTCATGGTGTGAAGCGTCGAGCCGACTAGGGAACCTCTTCCCGCCAGGATGGA 181  
QY 182 GTCGCATCAGTCGCGCGCTATTGCGCGGCTGTTCTTCCCTGTGT 226  
DB 182 GTCGCATCAGTCGCGCGCTATTGCGCGGCTGTTCTTCCCTGTGT 226

## RESULT 2

US-09-513-999C-33878  
; Sequence 33878, Application US/09513999C  
; Patent No. 6783961  
; GENERAL INFORMATION:  
; APPLICANT: Dumas Milne Edwards, J.B.  
; APPLICANT: Duclert, A.  
; APPLICANT: Giordano, J.Y.  
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.  
; FILE REFERENCE: 59.US2.REG  
; CURRENT APPLICATION NUMBER: US/09/513,999C  
; CURRENT FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/122,487  
; PRIOR FILING DATE: 1999-02-26  
; NUMBER OF SEQ ID NOS: 36681  
; SOFTWARE: Patent.pm  
; SEQ ID NO 33878  
; LENGTH: 218  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-513-999C-33878

Query Match 7.4%; Score 218; DB 4; Length 218;  
Best Local Similarity 100.0%; Pred. No. 4.5e-56;  
Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1499 ATAGAAATTATTACTACAGAGGTTCCACTTGTGTTTGAAGAAGAGCCCTTCTGAGGGC 1558  
DB 1 ATAGAAATTATTACTACAGAGGTTCCACTTGTGTTTGAAGAAGAGCCCTTCTGAGGGC 60  
QY 1559 TGTCCATCTAGTGAGTACCTATGGAAGAAGAGCGCTGAGAGGGCCGAATAAGTGTAGTT 1618  
DB 61 TGTCCATCTAGTGAGTACCTATGGAAGAAGAGCGCTGAGAGGGCCGAATAAGTGTAGTT 120  
QY 1619 GATTACTCATACCTAGAAAGTGACCTTCCCATTTCTGCCAGACCAGCTTGTAGTAACAA 1678  
DB 121 GATTACTCATACCTAGAAAGTGACCTTCCCATTTCTGCCAGACCAGCTTGTAGTAACAA 180  
QY 1679 CTGATAGATTATATTTTGGAGGTGCATCCAGTGACCT 1716  
DB 181 CTGATAGATTATATTTTGGAGGTGCATCCAGTGACCT 218

## RESULT 3

US-09-513-999C-11013  
; Sequence 11013, Application US/09513999C  
; Patent No. 6783961  
; GENERAL INFORMATION:  
; APPLICANT: Dumas Milne Edwards, J.B.  
; APPLICANT: Duclert, A.  
; APPLICANT: Giordano, J.Y.  
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.  
; FILE REFERENCE: 59.US2.REG  
; CURRENT APPLICATION NUMBER: US/09/513,999C  
; CURRENT FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/122,487  
; PRIOR FILING DATE: 1999-02-26  
; NUMBER OF SEQ ID NOS: 36681  
; SOFTWARE: Patent.pm  
; SEQ ID NO 11013  
; LENGTH: 195  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-513-999C-11013

Query Match 6.6%; Score 195; DB 4; Length 195;  
Best Local Similarity 100.0%; Pred. No. 4.5e-49;  
Matches 195; Conservative 0; Mismatches 0; Indels 0; Gaps 8;

QY 1 ATTTTGGGCTTTCGCTTCCACCGCAGCCGCGCTACCCAGTCTTCCGGTATCGCGTTG 60  
DB 1 ATTTTGGGCTTTCGCTTCCACCGCAGCCGCGCTACCCAGTCTTCCGGTATCGCGTTG 60  
QY 61 CTCAGGGGCTTTTCAACCTCTGTAGTCGGAACCATCGCGAGCGCGTGGGGGACT 120  
DB 61 CTCAGGGGCTTTTCAACCTCTGTAGTCGGAACCATCGCGAGCGCGTGGGGGACT 120  
QY 121 CCTATCCATGCTGTGAAGCGTCGAGCCGACTAGGGAACCTTCTTCCCGCCAGGATGGA 180  
DB 121 CCTATCCATGCTGTGAAGCGTCGAGCCGACTAGGGAACCTTCTTCCCGCCAGGATGGA 180  
QY 181 AGTCGCATCAGTCGC 195  
DB 181 AGTCGCATCAGTCGC 195

## RESULT 4

US-08-232-463-14/c  
; Sequence 14, Application US/08232463  
; Patent No. 5670367  
; GENERAL INFORMATION:  
; APPLICANT: DORNER, F.  
; APPLICANT: SCHEIFLINGER, F.  
; APPLICANT: FALKNER, F. G.

```

; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-Fls
; US-08-232-463-14

```

```

Query Match 1.7%, Score 51; DB 1; Length 7218;
Best Local Similarity 3.7%; Pred. No. 0.00076;
Matches 12; Conservative 189; Mismatches 124; Indels 0; Gaps 0;

Qy 2228 AAGTGCAGCTGTTGGGGAGCCAGAAAGTGAATGTCAGACTCGGTACAGCGTGACGTT 2287
Db 1370 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1311

Qy 2288 CTTTCTCGAGGAGACACACATGTCAAAGAAAGGTAACCTTCTCTTGAAGAAAGTT 2347
Db 1310 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1251

Qy 2348 ACTGAGTATTATATAAGTGTGATGAGGATCGCAAGACCATCGGAGAAATTTGCAAG 2407
Db 1250 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1191

Qy 2408 GATGATGCAGGTTCCAGAAACGAATTCAGAAACAGAGATGCTATTGGATATGCTTG 2467
Db 1190 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1131

Qy 2468 ACATTGACACAGAAAGAAATGTTTAATAGACTCCAGGAAACATGCTTCAAGGACTT 2527
Db 1130 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1071

Qy 2528 AATGTTCTCAAGCAATGTTGAGTTG 2552
Db 1070 RRRATCGCAAGCTCCCTCGACCTG 1046

```

```

RESULT 5
US-09-806-708B-22/c
; Sequence 22, Application US/09806708B

```

```

; Patent No. 6784342
; GENERAL INFORMATION:
; APPLICANT: The University of British Columbia
; TITLE OF INVENTION: Regulation of Embryonic Transcription in Plants
; FILE REFERENCE: 4810-58741
; CURRENT APPLICATION NUMBER: US/09/806,708B
; CURRENT FILING DATE: 2001-04-03
; PRIOR APPLICATION NUMBER: US 60/147,133
; PRIOR FILING DATE: 1999-08-04
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 22
; LENGTH: 1141
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; NAME/KEY: promoter
; LOCATION: (1)..(1141)
; OTHER INFORMATION: consensus sequence of A.t., L.a., and B.n. FAE1 promoters
; US-09-806-708B-22

```

```

Query Match 1.5%, Score 45.4; DB 4; Length 1141;
Best Local Similarity 11.6%; Pred. No. 0.01;
Matches 109; Conservative 325; Mismatches 498; Indels 9; Gaps 2;

Qy 1773 GTTTGATAGTAGCTCTGTCTCAGACTCAGACCTTGAACAGACCTTGAAGGGCTTC 1832
Db 1094 GTAWTTHAKRGATMCWYWTCTGTRRRCWRTYAMRTWYTRSNANWSCATKBMWWTMKWY 1035

Qy 1833 ACCTTGGAACTCTTTCTGCGAGTGTAGATCTTATATATATATATATATATATATAT 1892
Db 1034 ATKYRTAWYAMWCWNRNNNNNNCATNGYAKSCATNNAMVATTTRWAAAYAAAKWARWAGNNM 975

Qy 1893 TTCAGACTGCTGCCAGAAATTTCTCTGGAAGAGCCTTCTGATTCAGAGAGGATTTGCTG 1952
Db 974 RMYGAAAGNKGWCMGAATWATGBWADTAGMNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 915

Qy 1953 GCAAGTCTGATCTAGAGAAATCTCTCCAGTCTGGAAGCCTTCTCTGAGACCCCTGAGCATA 2012
Db 914 TACYNRAATNNKWAATHMKWTHGHSKRRTTRHTRTCRRTKYNNNNNNNNNNNNNNNNNNNN 855

Qy 2013 GTTCTGGGAGGAAGATGACTGGGAATCTAGTCAGATGAAGCAGAGAGTCTCAAACTGT 2072
Db 854 RW-----MNAWMTTRNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 802

Qy 2073 GGAACCTATTCTGTAATCTGATGACCCCTACAACTTTTAAATTTTAAAGGCTCTTTTC 2132
Db 801 CHYTTANABBCYRANNNNNAARWARTCNNTFMHAAVTTTHTDWCYKTWMTWYDMMTTMB 742

Qy 2133 AAACATCAGGGGAAATAGAGAAAGGCTGTGCTGACTCAAAGACCCCATCTGAGTCCATTG 2192
Db 741 TTTTRNNMTTSTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 682

Qy 2193 TGCCCATTTCTGAGTGTACACCTTACTTCTTGTAGGTGAGCTGTTGGGAGGCCAAG 2252
Db 681 VMERRRMTNTKRWYSTRHHYTGATNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 622

Qy 2253 AAAGTGAATGTCAGACTCGGTACAGCGTACCTTCTTTCTGAGGAGAGACACACATG 2312
Db 621 DGMVTRKKVKRRDTCCTYVDVWADSWWWTANMRCRDVTYTRNNYCKSYAHSYHWSN 562

Qy 2313 TCAAAAGAAAAAGGTAACCTTCTTGAAGAAAGTTACTAGTATTATATAAGTGGTGATG 2372
Db 561 NAMWYRYSARNWSSMARWTRNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 502

Qy 2373 AGGATCGCAAGACCATCGGAGAAAGATTTGCAAGGATGATGATGATGATGATGATGATG 2432
Db 501 ARETTTVDSCNCAKSMWGRNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 442

Qy 2433 TTCAAGAAACAGAGATGCTATTGGATATTGCTTGTGACATTGTAACACAGAGAAAGATGT 2492
Db 441 NMAWCRRAYCCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 384

```





Db 189 WMKGSRRRATSRYGMMSSMYGASGRMSSMCSASTRMSSASCMYMMWSAGSYASCAWM 248  
Qy 581 CGGCTGAGTTACTGAGCAAACTCTCTCCAGCTCTCTGCGCGCTCCCGGATTGCTT 640  
Db 249 SKYRCAKWSTSYWYWRASWKKSKYCAWSRKGSKCCWYSRKSGSKYTCWGGSCCCGCG 308  
Qy 641 CAGAGGTGCTAAATTGGAGCC 662  
Db 309 CAGCAGCAGCAGCTGCCAGC 330

## RESULT 9

US-09-621-976-8976/c  
; Sequence 8976, Application US/09621976  
; Patent No. 6639063  
; GENERAL INFORMATION:  
; APPLICANT: Dumas Milne Edwards, J.B.  
; APPLICANT: Jobert, S.  
; APPLICANT: Giordano, J.Y.  
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.  
; FILE REFERENCE: GENSET.054PR2  
; CURRENT APPLICATION NUMBER: US/09/621,976  
; CURRENT FILING DATE: 2000-07-21  
; NUMBER OF SEQ ID NOS: 19335  
; SOFTWARE: Patent.pm  
; SEQ ID NO 8976  
; LENGTH: 399  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-621-976-8976

Query Match 1.4%; Score 41.4; DB 4; Length 399;  
Best Local Similarity 19.5%; Pred. No. 0.077;  
Matches 56; Conservative 114; Mismatches 110; Indels 7; Gaps 1;  
Qy 2401 TGAAGGATGGATGCGAGGTTCCAGAAACGAATTCAGAAACAGAGATGCTATTGGATA 2460  
Db 391 TGAAGGTTGGAGCTATGAGCATCAATCAGCCAAAGKMWKWSGAMKWRGTGGWYMYM 332  
Qy 2461 TTGCTTGACATTTGAACACAGAGAAGNATGTTTAAAGACTCCAGGGAACATGCTCAA 2520  
Db 331 TSGSYRRTYTSAMGGRAMKRRKTKTKMYWMKGKGGKGGSTYAMRSRGTGRWSYR 272  
Qy 2521 AGGACTTAATGTTCTCAAGCAATGTTGAGTTGGAGCTGTAGTCTAGCTAGCATACAC 2580  
Db 271 AMRGSKSWGGSYRAGYRSSRWSYSAWRRKKMTCKGRSSGSRSTGYAWYK 212  
Qy 2581 TACCTCTTACCTGAGAGGTGCTTTTAAACAAATCTTGGCAGCTGCTCTTTGACATTT 2640  
Db 211 KSWCTSRKMYKKRRKKRRKCTST-----KRTCYRSGSTYKWKAYYTKRRKRWTR 159  
Qy 2641 TTTTITTAGAGGAAATGTAACCTTGGATCTAGTTTAAATTTTTTTTTT 2687  
Db 158 TTYYYKSYMSMKTKRMKTAYYTKRWKMTRTKWTCTMCKKCTY 112

## RESULT 10

US-09-806-708B-22  
; Sequence 22, Application US/09806708B  
; Patent No. 6784342  
; GENERAL INFORMATION:  
; APPLICANT: The University of British Columbia  
; TITLE OF INVENTION: Regulation of Embryonic Transcription in Plants  
; FILE REFERENCE: 4810-58741  
; CURRENT APPLICATION NUMBER: US/09/806,708B  
; PRIOR FILING DATE: 2001-04-03  
; CURRENT FILING DATE: 2001-04-03  
; PRIOR FILING DATE: 1999-08-04  
; NUMBER OF SEQ ID NOS: 23  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 22  
; LENGTH: 1141

; TYPE: DNA  
; ORGANISM: Artificial sequence  
; FEATURE:  
; NAME/KEY: promoter  
; LOCATION: (1)..(1141)  
; OTHER INFORMATION: consensus sequence of A.t., L.a., and B.n. FAE1 promoters  
US-09-806-708B-22

Query Match 1.4%; Score 40.8; DB 4; Length 1141;  
Best Local Similarity 9.0%; Pred. No. 0.26;  
Matches 65; Conservative 291; Mismatches 363; Indels 4; Gaps 1;  
Qy 2175 CCCCATCTGAGTCATTTGCGCCATTTCTGAGTGTCACACCTTACTTTCTTGTGAAGTGC 2234  
Db 53 MSKSRKWTWARMYCKYRRWYNNKSRWKGWYKKYKKBANNTSBRVHARRWKMKTAYBM 112  
Qy 2235 AGCTGTGGGAGCCCAAGAGTGAATGTCACAGCTCGGTACAGCGTGACGTTCTTCTG 2294  
Db 113 TMTNKGKGTGWRHRYWRWAMBTVDDHYVYVANNNAWTTWCMMDKDKRTRWKKNNNA 172  
Qy 2295 GAGGAAGACACACACATGTCAAAAGAAAGGTAACCTTCTTGAAGAACTTACTGAGT 2354  
Db 173 TGWDDDTKYHMMWNNNGCBTVWVRYKTDWDWSBKRMNYGMBWKNWSYDVYVWVWDD 232  
Qy 2355 ATTATATAAGTGTGATGAGGATCGCAAGGACCATCGGGAAGAAATTTGCAAGGATGAT 2414  
Db 233 MCKRVRWVTRGRMRNTYMWABTAHRRRYNNGWTBAMAYRRTWMTNNNNNNKAMCKRA 292  
Qy 2415 GCAGGTTCCAGAAACGAATTCAGAAACAGAAAGATGCTATTGATGATTGCTGACATTTG 2474  
Db 293 KYGMWNRABVNSTCTTWKSKTKRTYSCWANNCRAGDANKDKHWKWSAAMGVYVNNNN 352  
Qy 2475 AACACAGAGAAAGATGTTTAAATAGACTCCAGGGAACATGCTTCAAGAGCTTAATATTC 2534  
Db 353 NNWYKKARHBAWDVWHSWKWANAHAHYSRKWTBYKRTVNNNNNGTTWKKEMWA 412  
Qy 2535 TCAAGCAATGTTGAGTTGGAGCTGTAGTCTAGCTAGCATACACTCTTACTGA 2594  
Db 413 WYKMDMDWBGTYNNNNNGRTYTGWTKKMWYTKWKANNCKWRAWDHKTCTHNNTTW 472  
Qy 2595 GAGGTGCTTTTAAA-----AACAAATCTTGGCAGCTGCTCTTGGACATTTTTTTT 2650  
Db 473 WKKTYYNNCYKSMNTNGSKSHRBAAYVYTWMMWRYAHANNNNNDYMWKACTWYKYBV 532  
Qy 2651 AGGAATTAACCTGAGTCTAGTTTAAATTTTTTTTTCACACATATCCCATCAGAAAC 2710  
Db 533 CSKWNYYAAWYTKSSWNTYTSRYRWKTNNSWRSWSDTRSMGRANNYARABHYGYKNTR 592  
Qy 2711 ATTGAGTTTGAAGCCAGCCTGATTAATGAAGATGAACCTAGTGTGATTTCTTAATCTCC 2770  
Db 593 WBSWHTWBHBRAGAAHYWMBMYBAKCHCMKAWYKAKYAGAGGAGGNNNNNNNNNN 652  
Qy 2771 CTTTTTGTAGTTAGTGTGCTTTTAAATGCTCTTTGCTGTCATGAGTGAAGGG 2830  
Db 653 NNATCARDYYAASRYAMANKWYKYYKBAANNAYYTHANNWGWGWNNA'FDTRTWKN 712  
Qy 2831 GACCTTTTGTAGTTGCTTTTTCACCTTTCACCTTTCACCTTTCACCTTTCACCTTTC 2890  
Db 713 NNNNAGTWKNNNNNNNAKNAKNAKNAKNAKNAKNAKNAKNAKNAKNAKNAKNAKNAK 772  
Qy 2891 TAG 2893  
Db 773 NNG 775

## RESULT 11

US-09-949-016-15525  
; Sequence 15525, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF





**THIS PAGE BLANK (USPTO)**

Result No.	Score	Query #		DB	ID	Description
		Match	Length			
1	2942	100.0	2942	19	US-10-650-482-1	Sequence 1, Appl
2	2185.8	74.3	2324	9	US-09-925-299-168	Sequence 168, Appl
3	2185.8	74.3	2324	10	US-09-935-239-168	Sequence 168, Appl
4	1324	45.0	5468	18	US-10-650-482-3	Sequence 3, Appl
5	828.6	28.2	8883	18	US-10-236-115-224	Sequence 224, Appl
6	435.8	14.8	465	22	US-10-756-149-3997	Sequence 3997, Ap
7	265.8	9.0	558	9	US-09-878-178-2040	Sequence 2040, Ap





```
Qy 1441 GTTTGTTCTCTGCTGAGACATTTCTTGGAACACCCAGGAATCCACTGAAGAAAAAT 1500
Db 1441 GTTTGTTCTCTGCTGAGACATTTCTTGGAACACCCAGGAATCCACTGAAGAAAAAT 1500
Qy 1501 AGAATTATTAACCTACAGAGTTTCCACTGCTTTTGGAGAAGAGAGCCCTTCTGAGGCTG 1560
Db 1501 AGAATTATTAACCTACAGAGTTTCCACTGCTTTTGGAGAAGAGAGCCCTTCTGAGGCTG 1560
Qy 1561 TCCATCTAGTCAGATACCTATGGAAGAGAGCCCTGAGAGGCGCGAATAAGTGTAGTTGA 1620
Db 1561 TCCATCTAGTCAGATACCTATGGAAGAGAGCCCTGAGAGGCGCGAATAAGTGTAGTTGA 1620
Qy 1621 TTACTCATPACCTAGAAGGTGACCTTCCCATTTCTGCCAGACCCAGCTTGTAGTAACAAACT 1680
Db 1621 TTACTCATPACCTAGAAGGTGACCTTCCCATTTCTGCCAGACCCAGCTTGTAGTAACAAACT 1680
Qy 1681 GATAGATTATATTTTGGAGGTGATCCAGTGACCTGGAAACAAGTTCTGATCAGAAGG 1740
Db 1681 GATAGATTATATTTTGGAGGTGATCCAGTGACCTGGAAACAAGTTCTGATCAGAAGG 1740
Qy 1741 TGAGGATGGGATCAGGAAGCTGAGGATGATGTTTGTAGTAGTACACTGTCAGA 1800
Db 1741 TGAGGATGGGATCAGGAAGCTGAGGATGATGTTTGTAGTAGTACACTGTCAGA 1800
Qy 1801 CTCAGACCTTGAACAAGACCTTGAAGGCTTCACTTTTGGAACTCTTTCTGCAGTGTAGA 1860
Db 1801 CTCAGACCTTGAACAAGACCTTGAAGGCTTCACTTTTGGAACTCTTTCTGCAGTGTAGA 1860
Qy 1861 TCCTTATTAATCCCAAGACTTTTACAGCAAAATTCAGACTGCTGCCAAGATTGTTCTCTGA 1920
Db 1861 TCCTTATTAATCCCAAGACTTTTACAGCAAAATTCAGACTGCTGCCAAGATTGTTCTCTGA 1920
Qy 1921 AGAGCCTTCTGATTCAGAGAAGGATTTGCTGCGCAAGTCTGATCTAGAGAATTCCTCCCA 1980
Db 1921 AGAGCCTTCTGATTCAGAGAAGGATTTGCTGCGCAAGTCTGATCTAGAGAATTCCTCCCA 1980
Qy 1981 GTCTGGAAGCCTTCTCTGAGACCCCTGAGCATAGTTCTGGGGAGGAAGATGACTGGGAATC 2040
Db 1981 GTCTGGAAGCCTTCTCTGAGACCCCTGAGCATAGTTCTGGGGAGGAAGATGACTGGGAATC 2040
Qy 2041 TAGTGCAGATGAAGCAGAGAGTCTCAAACTGTGGAATCATCTGTAACTTGATGACCC 2100
Db 2041 TAGTGCAGATGAAGCAGAGAGTCTCAAACTGTGGAATCATCTGTAACTTGATGACCC 2100
Qy 2101 CTACAACCTTTAAATTTTAAAGGCTCTCTTTTCAAACATCAGGGGAAAATGAGAAGGCTG 2160
Db 2101 CTACAACCTTTAAATTTTAAAGGCTCTCTTTTCAAACATCAGGGGAAAATGAGAAGGCTG 2160
Qy 2161 TCGTGACTCAAAGACCCCATCTGAGTCCATTTGTGGCCATTTCTGAGTGTCAACCTTTACT 2220
Db 2161 TCGTGACTCAAAGACCCCATCTGAGTCCATTTGTGGCCATTTCTGAGTGTCAACCTTTACT 2220
Qy 2221 TTCTTTGAAGTGACGCTGTTGGGAGCCAAAGAAAGTGAATGTCCAGACTCGGTACAGCG 2280
Db 2221 TTCTTTGAAGTGACGCTGTTGGGAGCCAAAGAAAGTGAATGTCCAGACTCGGTACAGCG 2280
Qy 2281 TGAGCTTCTTTCTGAGGAGACACACACATGTCCAAAGAAAAGGTAACTTCTCTTGA 2340
Db 2281 TGAGCTTCTTTCTGAGGAGACACACACATGTCCAAAGAAAAGGTAACTTCTCTTGA 2340
Qy 2341 AGAAGTTACTGAGTATTTATATAAGTGGTGAAGGATCGCAAGGACCATGGGAAGAAAT 2400
Db 2341 AGAAGTTACTGAGTATTTATATAAGTGGTGAAGGATCGCAAGGACCATGGGAAGAAAT 2400
Qy 2401 TGCAGGGATGATGTCAGGTTCCAGAAACGAATTTCAAGAAACAGAAATGCTATTGGATA 2460
Db 2401 TGCAGGGATGATGTCAGGTTCCAGAAACGAATTTCAAGAAACAGAAATGCTATTGGATA 2460
Qy 2461 TTGCTTGACATTTGAAACACAGAGAAGAAATGTTTAAATAGACTCCAGGGGAACATGCTTCAA 2520
Db 2461 TTGCTTGACATTTGAAACACAGAGAAGAAATGTTTAAATAGACTCCAGGGGAACATGCTTCAA 2520
```

```
Qy 2521 AGGACTTAATGTTCTCAAGCAATGTTGAGTTGGCAGCCTGTAGTCTTAGCTAGCATACAC 2580
Db 2521 AGGACTTAATGTTCTCAAGCAATGTTGAGTTGGCAGCCTGTAGTCTTAGCTAGCATACAC 2580
Qy 2581 TACCTCTTACCTCAGAGGTGCTCTTTTAAAAAACAAATCTTGGCAGCTGTCTCTTTGACATTT 2640
Db 2581 TACCTCTTACCTCAGAGGTGCTCTTTTAAAAAACAAATCTTGGCAGCTGTCTCTTTGACATTT 2640
Qy 2641 TTTTCTTTAGAGGAAATGTAACCTTGGATCTAGTTAAATTTTGTGCAACATATCCC 2700
Db 2641 TTTTCTTTAGAGGAAATGTAACCTTGGATCTAGTTAAATTTTGTGCAACATATCCC 2700
Qy 2701 ACTCAGAAACATTCAGGTTTGAAGCCAGCCCTGATTAATGAAGGATGAATAGTGTGATTT 2760
Db 2701 ACTCAGAAACATTCAGGTTTGAAGCCAGCCCTGATTAATGAAGGATGAATAGTGTGATTT 2760
Qy 2761 CTAACTCTCCCTTTTGTATTTAGTTGGATGCTCTTTTAAATCTCTTTCSCCTGCAATGAG 2820
Db 2761 CTAACTCTCCCTTTTGTATTTAGTTGGATGCTCTTTTAAATCTCTTTCSCCTGCAATGAG 2820
Qy 2821 GTGGAAGGGGACCTTTTGTAGTTGTCAATTTTGCACCTTTTCAAAACTTATTTTCTTGGAAA 2880
Db 2821 GTGGAAGGGGACCTTTTGTAGTTGTCAATTTTGCACCTTTTCAAAACTTATTTTCTTGGAAA 2880
Qy 2881 ACAATATTTATAGGGCTTAAAGCCCATTTTTCATTTCTTAATCTAAATATGTTGTGCCATC 2940
Db 2881 ACAATATTTATAGGGCTTAAAGCCCATTTTTCATTTCTTAATCTAAATATGTTGTGCCATC 2940
Qy 2941 TG 2942
Db 2941 TG 2942
```

## RESULT 2

```
US-09-925-299-168
; Sequence 168, Application US/09925299
; Patent No. US20020055627A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 168
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-299-168
```

```
Query Match 74.3%; Score 2185.8; DB 9; Length 2324;
Best Local Similarity 97.9%; Pred. No. 0;
Matches 2249; Conservative 6; Mismatches 36; Indels 6; Gaps 4;

Qy 456 GCTTCCGGTCTCGCCACCCTTTTCCCTCGGCGATCGCAAGCAGGCTCTTCTAAGTTCC 515
Db 32 GCTCAGGGGCTTTTCAACCCCTCTGTCACTCGGCGATCGCAAGCAGGCTCTTCTAAGTTCC 91
Qy 516 CGACGCTCTTGGCCCGGAAAACCTCGGGAAACCCCACTGCTTTTCTCTGCCCAGCCCCG 575
Db 92 CGACGCTCTTGGCCCGGAAAACCTCGGGAAACCCCACTGCTTTTCTCTGCCCAGCCCCG 151
Qy 576 AGACTCGGGTCAGTTACTGGAAGAAACTGCTCTCCAGGCTCTTGGCCGCGCTCCCCGGAT 635
Db 152 AGACTCGGGTCAGTTACTGGAAGAAACTGCTCTCCAGGCTCTTGGCCGCGCTCCCCGGAT 211
Qy 636 TGCTTCAGAAAGGTGCTAAATTTGAGGCCAACTTTTCGGTGGAAATGTTTCCGACCATGGC 695
```

Db 212 TGCTTCAGAGGTGCTAAATTTGGAGCCAACTTTTCGGTGGAAATGTTTCCGACAGATGGC 271  
 Qy 696 TAGATTTTGTCTGAGTCTACAGGCCCTGAGAGCCCTGAAGGACGGGAGAAACACAGCG 755  
 Db 272 TAGATTTTGTCTGAGTCTACAGGCCCTGAGAGCCCTGAAGGACGGGAGAAACACAGCG 331  
 Qy 756 CCCCCACAGCGCAGAAATCTTTGAGTTCGCTCGAGCTCGACTCCCTCAGACCCCTCGGTCA 815  
 Db 332 CCCCCACAGCGCAGAAATCTTTGAGTTCGCTCGAGCTCGACTCCCTCAGACCCCTCGGTCA 391  
 Qy 816 CAGTCCCTTGAATTTGGCTAGAGGAGGAGATCCATGCGCAATATCTGCGCCCCAGACCTAA 875  
 Db 392 CCAAGTCCCTTGAATTTGGCTAGAGGAGGAGATCCATGCGCAATATCTGCGCCCCAGACCTAA 451  
 Qy 876 AATTGGAGCTTAAGCCCAAGGAGAGTCTTTGGAGCCCTGCAGCAGAGGCTTTCTCTTAG 935  
 Db 452 AATTGGAGCTTAAGCCCAAGGAGAGTCTTTGGAGCCCTGCAGCAGAGGCTTTCTCTTAG 511  
 Qy 936 AGCAGCAGCTGTGGGAGTGGAGCTGTGGCCAGTAGCCTTCAATCCCGTCTGTACTCTA 995  
 Db 512 AGCAGCAGCTGTGGGAGTGGAGCTGTGGCCAGTAGCCTTCAATCCCGTCTGTACTCTA 571  
 Qy 996 ACCGGGAATCTGGCTCTTGGCCCTCTGGCCCTCTAAACATTCACGCAATAGCAATTTCA 1055  
 Db 572 ACCGGGAATCTGGCTCTTGGCCCTCTGGCCCTCTAAACATTCACGCAATAGCAATTTCA 631  
 Qy 1056 GTGTGGTATCTTATTTGCTGAACCTTCTACTGACTGCTTTCTAGGCTAGAGTCA 1115  
 Db 632 GTGTGGTATCTTATTTGCTGAACCTTCTACTGACTGCTTTCTAGGCTAGAGTCA 691  
 Qy 1116 GCTATCAGAACAGTGAATAGCGAGTGTAGTGGCTTCCAGACACTAACCCAGAGA 1175  
 Db 692 GCTATCAGAACAGTGAATAGCGAGTGTAGTGGCTTCCAGACACTAACCCAGAGA 751  
 Qy 1176 GAGCTGCTGAGAGAGGACCAATGTTCATCCAGCGCTGAGTGCAGAACTCATTTCCGG 1235  
 Db 752 GAGCTGCTGAGAGAGGACCAATGTTCATCCAGCGGTGARTGCAGAACTCATTTCCGG 811  
 Qy 1236 CTTGCTGGCAGGAGTGCACCTCTTCTACGGAAGGCTACAGAAATTCACCACTTTC 1295  
 Db 812 SCTGTGGCAGGAGTGCACCTCTTCTACGGAAGGCTACAGAAATTCACCACTTTC 871  
 Qy 1296 GCATGAACCGCTGGAATTCCTTCAACAGGCTAAACAGGGCAAGATTTACCCACCCCTG 1355  
 Db 872 GCATGAACCGCTGGAATTCCTTCAACAGGCTAGCAAGGGCAAGATTTACCCACCCCTG 931  
 Qy 1356 ACCAGATAATGGCTACCAAGCTGAGGAGGAAACAGCCCTTCTCGGATGGATCCAA 1415  
 Db 932 ACCAGATAATGGCTACCAAGCTGAGGAGGAAACAGCCCTTCTCGGATGGATCCAA 991  
 Qy 1416 AACACTGCAGAGATAACCCAAACACAGTTTGTCTGCTGGAGACATTCCTGGAAACA 1475  
 Db 992 AACACTGCAGAGATAACCCAAACACAGTTTGTCTGCTGGAGACATTCCTGGAAACA 1051  
 Qy 1476 CCAGGAATCCACTGAAGAAAAAATAGAAATTAATTAACATACAGAGTTCACCTTGTGG 1535  
 Db 1052 CCAGGAATCCACTGAAGAAAAAATAGAAATTAATTAACATACAGAGTTCACCTTGTGG 1111  
 Qy 1536 AAGAGAGAGCCCTTCTGAGGCTGTCCATCTAGTGAATACCTATGAAAGAGGCTG 1595  
 Db 1112 AAGAGAGAGCCCTTCTGAGGCTGTCCATCTAGTGAATACCTATGAAAGAGGCTG 1171  
 Qy 1596 GAGAGGCGGAATAAGTGTAGTTGATCTCATACTAGAGGTGACCTTCCCACTTCTG 1655  
 Db 1172 GAGAGGCGGAATAAGTGTAGTTGATCTCATACTAGAGGTGACCTTCCCACTTCTG 1231  
 Qy 1656 CCAGACCAAGCTTGTATGAACAACTGATATTAATTTGGAGGATGATCCAGTGACC 1715  
 Db 1232 CCAGACCAAGCTTGTATGAACAACTGATATTAATTTGGAGGATGATCCAGTGACC 1291  
 Qy 1716 TGGAAACAAGTTCTGATCCAGAGGTGAGGATTTGGATGAGGAAGCTGAGGATGATGTT 1775  
 Db 1292 TGGAAACAAGTTCTGATCCAGAGGTGAGGATTTGGATGAGGAAGCTGAGGATGATGTT 1351

Qy 1776 TTGATAGTAGTAGTCTCACTGTGAGACTCAGACCTTGAACCAAGACCTCGAAGGGCTTCACC 1835  
 Db 1352 TTGATAGTAGTAGTCTCACTGTGAGACTCAGACCTTGAACCAAGACCTCGAAGGGCTTCACC 1411  
 Qy 1836 TTTGGAATCTTTCTGAGTGTAGATCCTTATAATATCCCAAGAACTTTTACAGCAAAATTC 1895  
 Db 1412 TTTGGAATCTTTCTGAGTGTAGATCCTTATAATATCCCAAGAACTTTTACAGCAAAATTC 1471  
 Qy 1896 AGACTGTGCCAGAAATTTGTTCTCGAAGAGCTTCTGATTCAGAGAAGATTTGCTGGCA 1955  
 Db 1472 AGACTGTGCCAGAAATTTGTTCTCGAAGAGCTTCTGATTCAGAGAAGATTTGCTGGCA 1531  
 Qy 1956 AGTCTGATCTAGAGAAATTTCTCCAGTCTGGAAGCCCTTCTGAGACCCCTGAGCATAGTT 2015  
 Db 1532 AGTCTGATCTAGAGAAATTTCTCCAGTCTGGAAGCCCTTCTGAGACCCCTGAGCATAGTT 1591  
 Qy 2016 CTGGGAGGAAGATGACTGGGAATCTAGTGCAGATGAAGCAGAGAGTCTCAAACTGTGGA 2075  
 Db 1592 CTGGGAGGAAGATGACTGGGAATCTAGTGCAGATGAAGCAGAGAGTCTCAAACTGTGGA 1651  
 Qy 2076 ACTCATTTCTGTAATTTCTGATGACCCCTTACACCCCTTTAAATTTTAAAGCTCTCTTTCAA 2135  
 Db 1652 ACTCATTTCTGTAATTTCTGATGACCCCTTACACCCCTTTAAATTTTAAAGCTCTCTTTCAA 1711  
 Qy 2136 CATCAGGGGAAAAATGAGAAAGGCTGTCTGACTCAAAAGACCCCATCTGAGTCCATTTGG 2195  
 Db 1712 CATCAGGGGAAAAATGAGAAAGGCTGTCTGACTCAAAAGACCCCATCTGAGTCCATTTGG 1771  
 Qy 2196 CCATTTCTGAGTGTCAACCTTACTTTCTGTAAGGTGACGCTGTTGGGAGGCAAGAA 2255  
 Db 1772 CCATTTCTGAGTGTCAACCTTACTTTCTGTAAGGTGACGCTGTTGGGAGGCAAGAA 1831  
 Qy 2256 GTGAACTGCCAGACTCGGTACAGGTGACCTTCTTCTGAGGAGGAGACACACATGTCA 2315  
 Db 1832 GTGAACTGCCAGACTCGGTACAGGTGACGCTTCTTCTGAGGAGGAGACACACATGTCA 1891  
 Qy 2316 AAAGAAAAAGGTAACTTCTTCAAGAAAGTACTGAGTATTATATAAGTGGTGATGAGG 2375  
 Db 1892 AAAGAAAAAGGTAACTTCTTCAAGAAAGTACTGAGTATTATATAAGTGGTGATGAGG 1951  
 Qy 2376 ATCCCAAGGACCATGGAAGAAATTTGCAAGGGATGGAATGCAAGTTCAGAAACGAATTC 2435  
 Db 1952 ATCCCAAGGACCATGGAAGAAATTTGCAAGGGATGGAATGCAAGTTCAGAAACGAATTC 2011  
 Qy 2436 AAGAAACAGAGAGTGTATTGGATATTGCTTGACATTTGAAACAGAGAAAGAAATGTTTA 2495  
 Db 2012 AAGAAACAGAGAGTGTATTGGATATTGCTTGACATTTGAAACAGAGAAAGAAATGTTTA 2071  
 Qy 2496 ATAGACTCCAGGGAAACATGCTTCAAGGACCTTAATGTTCTCAAGCAATGTTGAGTGGCA 2555  
 Db 2072 ATAGACTCCAGGGAAACATGCTTCAAGGACCTTAATGTTCTCAAGCAATGTTGAGTGGCA 2131  
 Qy 2556 GCCTGTAGTCTAGTAGCATACACTACCTTACCTGAGAGGTGTCTTTTAAAAACAAA 2615  
 Db 2132 GCCTGTAGTCTAGTAGCATACACTACCTTACCTGAGAGGTGTCTTTTAAAAACAAA 2191  
 Qy 2616 TCTTGGCAGCTGTCTTTGACATTTTTTTTTTTTATAGAGAAATGTAACCTTGATCTAGTTT 2675  
 Db 2192 TCTTGGCAGCTGTCTTTGACATTTTTTTTTTTTATAGAGAAATGTAACCTTGATCTAGTTT 2247  
 Qy 2676 AATTTTTTTTTTGTCAACATATCCCACTCAGAAACATT-CAGGTTTGAAGCCAGCCCT-G 2733  
 Db 2248 TTTTTTTTTTTTGTCAACATATCCCACTCAGAAACATTCCAGGTTTGAAGCCAGCCCTGG 2307  
 Qy 2734 ATAATGAAGGATGAAT 2750  
 Db 2308 ATAATGAAGGATGAAT 2324



```

QY 2316 AAAAGAAAAAGTAACTTCTTGAAGAAGTTACTGAGTATTATATATAGTGGTATGAGG 2375
Db 1892 AAAGAAAAAAGTAACTTCTTGAAGAAGTTACTGAGTATTATATATAGTGGTATGAGG 1951
QY 2376 ATCGCAAAAGGACCATGGGAAGAAATTTGCAAGGGATGGATGCAAGGTTCAGAGAAACGAATTC 2435
Db 1952 ATCGCAAAAGGACCATGGGAAGAAATTTGCAAGGGATGGATGCAAGGTTCAGAGAAACGAATTC 2011
QY 2436 AAGAAACAGAGATGCTATTGGATATTGCTTGAACATTTGAAACACAGAGAAAGAAATGTTTA 2495
Db 2012 AAGAAACAGAGATGCTATTGGATATTGCTTGAACATTTGAAACACAGAGAAAGAAATGTTTA 2071
QY 2496 ATAGACTCCAGGACATGCTTCAAGACATTAATGTTCTCAAGCAATGTTGAGTTGSCA 2555
Db 2072 ATAGACTCCAGGACATGCTTCAAGACATTAATGTTCTCAAGCAATGTTGAGTTGSCA 2131
QY 2556 GCCTGAGTCTAGTAGCATACACTACTCTTACCTGAGAGGTGCTCTTTTAAACAA 2615
Db 2132 GCCTGAGTCTAGTAGCATACACTACTCTTACCTGAGAGGTGCTCTTTTAAACAA 2191
QY 2616 TCTGGAGAGTCTCTTGAACATTTTCTTATAGAGAAATGTAATGATCTAGTTT 2675
Db 2192 TCTGGAGAGTCTCTTGAACATTTTCTTATAGAGAAATGTAATGATCTAGTTT -GSMCTRGTA 2247
QY 2676 AATTTTTTTTTTGTCAACATATCCCACTCAGAAACAT -CAGGTTTGAAGCCAGCCCT-G 2733
Db 2248 TTTTTTTTTTTTGTCAACATATCCCACTCAGAAACATTTCCAGGTTTGAAGCCAGCCCTGG 2307
QY 2734 ATAATGAAGGATGAAT 2750
Db 2308 ATAATGAAGGATGAAT 2324

RESULT 4
US-10-650-482-3
; Sequence 3, Application US/10650482
; Publication No. US20040142345A1
; GENERAL INFORMATION:
; APPLICANT: Ron, David
; APPLICANT: Jousse, Celine
; TITLE OF INVENTION: METHODS OF SCREENING TEST COMPOUNDS USING GADD34L, AN eIF2alpha-S
; FILE OF INVENTION: PHOSPHATASE SUBUNIT
; FILE REFERENCE: 5986/11712-US1
; CURRENT APPLICATION NUMBER: US/10/650,482
; CURRENT FILING DATE: 2003-08-28
; PRIOR APPLICATION NUMBER: US 60/408,679
; PRIOR FILING DATE: 2002-09-06
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 5468
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-650-482-3

Query Match 45.0%; Score 1324; DB 19; Length 5468;
Best Local Similarity 72.5%; Pred. No. 0;
Matches 1964; Conservative 0; Mismatches 660; Indels 84; Gaps 16;

QY 96 CCATCGCGAGCCCGTGGGGGACTCTTATCCATGGTGTGAAGCGTCGAGCCGACTAGG 155
Db 151 CGTCCGCGCGCGGAGGGGGTCTTCTATGTTGAGGCGATCTCACAGCGGCTAGG 210
QY 156 GAACCTCTCTCC -CGCAGAGATGAAAGTCCGATCAGTCGCGGCTATTGCGCGGGCTGT 214
Db 211 AGCTCTCTCTCCAGCGGGATGACCTTAAACCGCGTCCGACCGCTTGCAGCGGCTC 270
QY 215 TCTTCCCTGTGTCTCGCGCGCGCTCGCATTCGCTCTGCTGCTGCTGCTGCTGCTG 274
Db 271 TGGGCGGCTCGGCTGAGAGACTGTTGCGGAGCCGCGCTCTCTGGGCTCTCTGCGG 330
QY 275 CTCGAGATCGGCTTGGAGACGCGACCGCTCGGCGAAGGCGCGAGACTCTGTAGGCT 334

```

```

Db 331 CGCGGGAATCGGACTGAGTACCCACTCGTGGCTGGCAAGGCGGAGACTGTGTAGACC 390
QY 335 TCCTCCGAATCCCGTGCAGCTTCAGCGGCTGAGCGCGCGGCGCTTACTGAGAGACTGTG 394
Db 391 TCGGATCCAGCTCGCTGAGCGCGCTGAGCTCTGTCTCTCTCTGCTGAGAAAGCGCC 450
QY 395 AAGAAAAAGAGATGAGCGCGGACAGGGCGATCGCGGAAACCGCTTGGCCCTTCGGCGG 454
Db 451 AAG-GAAGAGATGAGAGACAGGAAACGCAAGGGCCGGAAGCGGCTTGGCCCTCGGCTG 509
QY 455 GGCTTCGGTTCGGCCACCCCTTTTCCCTCGGAGATCGCAAGCAGGCTCTTCTAAAGTTC 514
Db 510 GGCTTCCTGGTTCGGCTGCGC---TTCCTTCGGCGATCGCAAGCGCTGCTCTTCGAGTTC 566
QY 515 CCGAGCGCTTCGGCCCGGAAACTCGGGAACCCCACTGCTTCTTCTCTGCCCCAGGCC 574
Db 567 CCGCGCGCTTCCTCTCGACAAATTCGCGGAA-----CTCGCTCTGCCC 611
QY 575 GAGACTCGGCTCAGTTACTGGAAGAACTGCTCTCCAGCTCTTTCGCGCGCTCCCCGGA 634
Db 612 GAGGCTCGGACCAAGTACTGGAACAAATGCTTCTCAGCTCTTTCGCTGCTCCCTAGC 671
QY 635 TTGCTTCAGAAAGTGTCTAAATTTGAGCCAACTTTTCGGTGAATGTTTCCGACCAAGTGG 694
Db 672 CTATTCAGAAAGTGTCTGCTTTGGAGCCAGCTTTCCGGGCGCTGATTCCTTACCAGATGG 731
QY 695 CTAGATTTTGTGAGTCTACAGCGGCTGAGAGCCCTGAAGGACGGGAGAACCCAGCC 754
Db 732 CTAGATTTTTCGCGCAAGTTACAGCGGCTTTCGAGAGCTTCGAGAGACGGGAGGAATCTGAC 791
QY 755 GCCCCACAGCGCAGAAATCTTTGAGTT-CGCTGCAGCTCGACTCTCTC---AGACCCCTC 810
Db 792 GCTCCACGGTGCAGAGTCTCTGAGTTA CACTGGGCTGGAATCTTTCGCGAAGACTCGC 851
QY 811 GGTCAACAGTCCCCTTGATTTGGCTTAGAGAGGGGATCCACTGGCAATPACTGCCCCCAGA 870
Db 852 GTCGTGAGTACTCTTGCATTTGGCTTAGAGAGGGGACTCCAGTGGCAGTGTCTGCTCTCAGA 911
QY 871 CCTAAATTTGGAGCTTAAGCCCAAGGCAAGTGTCTTTCGACCCCTGCAGCAGACAGGCTTTCT 930
Db 912 -CTGGAAGTTAAACTCAAGGCCCAAGAAAGAGCTTTTAGACTCTGACGCGCCACTTTTCT 970
QY 931 CTTAGACAGCAGCTGTGGGGAGTGGAGCTGTTGCCCAGTAGCTTCAATCCCTCTGTGTA 990
Db 971 CTTGAGCAGCAGCTGTGGGGAGTGGAGTGTGCTGCCCAGTAGCTTCAAGCTGTGTCTAGT 1030
QY 991 CTCTAACCGGGAACCTTGGCTCTTTCGCCCTCTGCGGCTCTTAAACATTCACGCTAGACAA 1050
Db 1031 CTCCACCGGAACTTGACTCTTCTCATCTCTGCGGCTCTGAGCGCTTCAGAGCTTTAGGTAA 1090
QY 1051 TTTTCAGTGTGTATCTTATTTGCTGAACCTTCTCTTCTGAGCTGCTTCTCTAGGCTAGA 1110
Db 1091 TTTCAAGGTAGTTTCTTATCTCTGAAACCTTCTCTTCTGAGTACTTCTTCTCCAGTTAGG 1150
QY 1111 AGTCAGCTATCAGAAACAGTGTATGAAATAGCGAGTGTAGTGGCTTTCAGACACTAACCCC 1170
Db 1151 GCTGCGCTGTCAGAGCAGCTGAGGTGGCCAGTTTGTGGTTTCCGAAACACTAACCCC 1210
QY 1171 AGAGAGCAGTGTCTGAGAGAGGACCAATTTGTATCCCAGCGCTGAGTGCAGAACTCAT 1230
Db 1211 AGAGAGCTGTCTATCTTTCTGAAAGATGTTGTCTCATCTCTGCGGCTTTCGCGGAGAGATGTC 1270
QY 1231 TCCGCGCTTCGTGGCAGGAGTGTCCACCTCTTCTTACGGAAGGCTTACAGAAATTCACCA 1290
Db 1271 GGCACCGCTTCGAGAGAGTGTTCGCGCTCTCTCTACAGAGGCTTCGCGAAATTCACCA 1330
QY 1291 TCTTCGATGAAACCGCTGGAATTCCTTCAACAGGCTTAAACAGGGGCAAGATTTACCCAC 1350
Db 1331 CCTTCGTATGAAACCGCTAGAAATTCCT---CCAGGCTTAAACAGGGGCAAGATTTACCCAC 1387
QY 1351 CCTTCAGCAGGATTAATGGCTTACCACTGAGGAGGAGAAACAGAGCTTCTCCGGATGGA 1410

```



```

Db 301 CATAGTTCTGGGGGAAGATGACTGGGAATCTAGTGCAGATGAAGCAGAGAGTCTTCAA 360
Qy 2069 ---CTGTGGAATCATCTCTGTAA---TTCTGATGACCCCTCAACCCCTTTAAATTTTAAG 2122
Db 361 ACTGTGGGAATCTCATCTCTGTAAATCTGGATGGACCCCTACACCCCTTTAAATTTTAAG 420
Qy 2123 GCTCCTTTTCAACATCATCGGGGAAATGAGAAAGGCTGTCGTGACTCAAAAGACCCCATCT 2182
Db 421 GCTCCTTTTCAACATCATCGGGGAAATGAGAAAGGCTGTCGTGACTCAAAAGACCCCATCT 480
Qy 2183 GAGTCATTTGGCCCATTTCTGAGTGTACACCTTACTTTCTTGTAGAGTGAGCTGTG 2242
Db 481 GAGTCATTTGGCCCATTTCTGAGTGTACACCTTACTTTCTTGTAGAGTGAGCTGTG 540
Qy 2243 GGGAGCAAGAAAGTGAATGTCAGACTCGGTACAGCGTGAAGTCTTTCTGAGGAAAGA 2302
Db 541 GGGAGCAAGAAAGTGAATGTCAGACTCGGTACAGCGTGAAGTCTTTCTGAGGAAAGA 600
Qy 2303 CACACATGTCAAAGAAAAAGGTAACTTCTTCTTGAAGAAAGTCTGAGTATTATATA 2362
Db 601 CACACATGTCAAAGAAAAAGGTAACTTCTTCTTGAAGAAAGTCTGAGTATTATATA 660
Qy 2363 AGTGTGATGAGATCGCAAGGACCATGGGAAGAAATTTGCAAGGATGATGCAAGTTC 2422
Db 661 AGTGTGATGAGATCGCAAGGACCATGGGAAGAAATTTGCAAGGATGATGCAAGTTC 720
Qy 2423 CAGAAACGAATTCAGAAACAGAAAGATGCTATTGGATATTGCTTGCACATTTGAACACAGA 2482
Db 721 CAGAAACGAATTCAGAAACAGAAAGATGCTATTGGATATTGCTTGCACATTTGAACACAGA 780
Qy 2483 GAAAGAAATTTTAATAGACTCCAGGAAACATCTTCAAAGGACTTAATTTCTCAAGCAA 2542
Db 781 GAAAGAAATTTTAATAGACTCCAGGAAACATCTTCAAAGGACTTAATTTCTCAAGCAA 840
Qy 2543 TGTGAGTGGAGCTGAGTCTAGTCTAGTACATACACTACT 2585
Db 841 TGTGAGTGGAGCTGAGTCTAGTCTAGTACATACACTACT 883

RESULT 6
US-10-756-149-3997
; Sequence 3997, Application US/10756149
; Publication No. US20050181375A1
; GENERAL INFORMATION:
; APPLICANT: Aziz, Natasha
; APPLICANT: Zlotnik, Albert
; TITLE OF INVENTION: NOVEL METHODS OF SCREENING FOR MODULATORS OF METASTATIC CANCER
; FILE REFERENCE: file
; CURRENT APPLICATION NUMBER: US/10/756,149
; CURRENT FILING DATE: 2004-01-12
; NUMBER OF SEQ ID NOS: 5818
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 3997
; LENGTH: 465
; TYPE: DNA
; ORGANISM: Homo Sapiens
US-10-756-149-3997

Query Match 14.8%; Score 435.8; DB 22; Length 465;
Best Local Similarity 99.1%; Pred. No. 1.9e-119;
Matches 459; Conservative 0; Mismatches 2; Indels 2; Gaps 2;

Qy 721 CCTGAGAGCCCTGAAGGGAGCGGAGAAACCCAGCCGCCCCACAGCGCAGAAATCTTTCAG 780
Db 1 CCTGAGAGCCCTGAAGGGAGCGGAGAAACCCAGCCGCCCCACAGCGCAGAAATCTTTCAG 60
Qy 781 TTCGTGCGAGCTCGACTCTCTCAGACCCCTCGGTGACACAGTCCCTTTGATTTGGCTAGAGA 840
Db 61 TTCGTGCGAGCTCGACTCTCTCAGACCCCTCGGTGACACAGTCCCTTTGATTTGGCTAGAGA 120
Qy 841 GGGGATCCACTGGCAATACTCGCCCCCAGACCTTAAATTTGAGCTTAAGGCCAAGGGAAG 900

```

```

Db 121 GGGGATCCACTGGCAATATCTGCCGCCAGACCTAAAAATGGAGCTTAAGGCCAAGGGAAG 180
Qy 901 TGTCTTGGACCTCAGCAGACAGGCTTTTCTTTAGACAGCAGCTGTGGGAGTGGAGCT 960
Db 181 TGTCTTGGACCTCAGCAGACAGGCTTTTCTTTAGACAGCAGCTGTGGGAGTGGAGCT 240
Qy 961 GTTCCCCAGTAGCTTCAATCCCGTCTGTACTCTTAACCCGGGAACTTGGCTCTTCGCCCTC 1020
Db 241 GTTCCCCAGTAGCTTCAATCCCGTCTGTACTCTTAACCCGGGAACTTGGCTCTTCGCCCTC 300
Qy 1021 TGGGCTCTTAAACATTCACCGCATAGACAAATTTTCAGTGTGGTATCTTATTTGCTGAACCC 1080
Db 301 TGGGCTCTTAAACATTCACCGCATAGACAAATTTTCAGTGTGGTATCTTATTTGCTGAACCC 360
Qy 1081 TTCTTACTGAGCTGCTTTCTTAGGCTAGAAAGTCAAGTATCAG-AACAGTATGGAATA 1139
Db 361 TTCTTACTGAGCTGCTTTCTTAGGCTAGAAAGTCAAGTATCAGAAACAGTATGGAATA 420
Qy 1140 GCGAGGTAGTGGCTTCCAGACAC-TAAACCCCGAGAGCAGCT 1181
Db 421 GCGAGGTAGTGGCTTCCAGACACCTTAAACCCCGAGAGCAGCT 463

RESULT 7
US-09-878-178-2040
; Sequence 2040, Application US/09878178
; Patent No. US20020177552A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yugu
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.527
; CURRENT APPLICATION NUMBER: US/09/878,178
; CURRENT FILING DATE: 2001-06-08
; NUMBER OF SEQ ID NOS: 2237
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2040
; LENGTH: 558
; TYPE: DNA
; ORGANISM: Homo sapien
; NAME/KEY: misc feature
; LOCATION: (1)...(558)
; OTHER INFORMATION: n = A,T,C or G
US-09-878-178-2040

Query Match 9.0%; Score 265.8; DB 9; Length 558;
Best Local Similarity 96.4%; Pred. No. 2.5e-68;
Matches 270; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 2663 TTGGATCTAGTTTAAATTTTTCACACATATCCACTCAGAAACATTCAGGTTGA 2722
Db 3 TTGGGNNCTAGNTTAAATTTTTCACACATATCCACTCAGAAACATTCAGGTTGA 62
Qy 2723 AGCCAGCCCTGATAATGAAGATGAAGTGTGTGATTTCTTAATCCTCCCTTTTGTATTT 2782
Db 63 AGCCAGCCCTGATAATGAAGATGAAGTGTGTGATTTCTTAATCCTCCCTTTTGTATTT 122
Qy 2783 AGTTGGAATGCTTTTAAATGTCTTTGCTCTGATAGGTGGAAGGGGACCTTTTGTAG 2842
Db 123 AGTTGGAATGCTTTTAAATGTCTTTGCTCTGATAGGTGGAAGGGGACCTTTTGTAG 182
Qy 2843 TTGTCAATTTTGCATTTCAAACTTATTTTCTTGGAAAAAATATTTATAGGCTTAAAG 2902
Db 183 TTGTCAATTTTGCATTTCAAACTTATTTTCTTGGAAAAAATATTTATAGGCTTAAAG 242
Qy 2903 CCAATTTTCAATTTCTAATCTAAATTTATGTGTGCTATCTG 2942
Db 243 CCAATTTTCAATTTCTAATCTAATTTATGTGTGCTATCTG 282

```



```
RESULT 8
US-10-046-935-2040
; Sequence 2040, Application US/10046935
; Publication No. US20020156011A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yugu
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Secrist, Heather
; APPLICANT: Wang, Aijun
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.527C1
; CURRENT APPLICATION NUMBER: US/10/046.935
; CURRENT FILING DATE: 2002-01-15
; NUMBER OF SEQ ID NOS: 2239
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2040
; LENGTH: 558
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 8, 9, 14, 354, 527
; OTHER INFORMATION: n = A,T,C or G
US-10-046-935-2040

Query Match          9.0%; Score 265.8; DB 13; Length 558;
Best Local Similarity 96.4%; Pred. No. 2.5e-68;
Matches 270; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 2663 TTGGATCTAGTTAAATTTTTTTTTTGTGCAACATATCCCACTCAGAAACATTCAGGTTTGA 2722
Db 3 TTGGNNCTAGNTTAATTTTTTTTTTGTGCAACATATCCCACTCAGAAACATTCAGGTTTGA 62

Qy 2723 AGCCAGCCCTGATATGAAGATGAACATAGTGTGATTTCTAATCTCTCTTTTGTGATTT 2782
Db 63 AGCCAGCCCTGATATGAAGATGAACATAGTGTGATTTCTAATCTCTCTTTTGTGATTT 122

Qy 2783 AGTTGGATGTCTTTTAATGTCTTTTCCCTGCATGAGTGGAAAGGGGACCTTTTGTAG 2842
Db 123 AGTTGGATGTCTTTTAATGTCTTTTCCCTGCATGAGTGGAAAGGGGACCTTTTGTAG 182

Qy 2843 TTGTCAATTTTGCACCTTCAAACTTATTTTCTTGGAAACAAATATTTATAGGGCTTAAAG 2902
Db 183 TTGTCAATTTTGCACCTTCAAACTTATTTTCTTGGAAACAAATATTTATAGGGCTTAAAG 242

Qy 2903 CCCATTTTCAATCTAATCTAAATTAATGTGCTATCTG 2942
Db 243 CCCATTTTCAATCTAATCTAATTAATGTGCTATCTG 282

RESULT 9
US-10-046-502-2040
; Sequence 2040, Application US/10146502
; Publication No. US20030069180A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yugu
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Secrist, Heather
; APPLICANT: Wang, Aijun
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.527C2
; CURRENT APPLICATION NUMBER: US/10/146,502
; CURRENT FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 2241
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2040
; LENGTH: 558
; TYPE: DNA
; ORGANISM: Homo sapiens
```

```
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 8, 9, 14, 354, 527
; OTHER INFORMATION: n = A,T,C or G
US-10-146-502-2040

Query Match          9.0%; Score 265.8; DB 14; Length 558;
Best Local Similarity 96.4%; Pred. No. 2.5e-68;
Matches 270; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 2663 TTGGATCTAGTTAAATTTTTTTTTTGTGCAACATATCCCACTCAGAAACATTCAGGTTTGA 2722
Db 3 TTGGNNCTAGNTTAATTTTTTTTTTGTGCAACATATCCCACTCAGAAACATTCAGGTTTGA 62

Qy 2723 AGCCAGCCCTGATATGAAGATGAACATAGTGTGATTTCTAATCTCTCTTTTGTGATTT 2782
Db 63 AGCCAGCCCTGATATGAAGATGAACATAGTGTGATTTCTAATCTCTCTTTTGTGATTT 122

Qy 2783 AGTTGGATGTCTTTTAATGTCTTTTCCCTGCATGAGTGGAAAGGGGACCTTTTGTAG 2842
Db 123 AGTTGGATGTCTTTTAATGTCTTTTCCCTGCATGAGTGGAAAGGGGACCTTTTGTAG 182

Qy 2843 TTGTCAATTTTGCACCTTCAAACTTATTTTCTTGGAAACAAATATTTATAGGGCTTAAAG 2902
Db 183 TTGTCAATTTTGCACCTTCAAACTTATTTTCTTGGAAACAAATATTTATAGGGCTTAAAG 242

Qy 2903 CCCATTTTCAATCTAATCTAATTAATGTGCTATCTG 2942
Db 243 CCCATTTTCAATCTAATCTAATTAATGTGCTATCTG 282

RESULT 10
US-09-920-300A-45/c
; Sequence 45, Application US/09920300A
; Patent No. US20020136728A1
; GENERAL INFORMATION:
; APPLICANT: King, Gordon E.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Xu, Jiangchun
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.547
; CURRENT APPLICATION NUMBER: US/09/920,300A
; CURRENT FILING DATE: 2001-07-31
; NUMBER OF SEQ ID NOS: 1789
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 45
; LENGTH: 706
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 34, 35, 39
; OTHER INFORMATION: n = A,T,C or G
US-09-920-300A-45

Query Match          6.8%; Score 199; DB 9; Length 706;
Best Local Similarity 94.8%; Pred. No. 3.4e-48;
Matches 217; Conservative 0; Mismatches 10; Indels 2; Gaps 1;

Qy 2716 GGTTTGAAGCCAGCCCTGATAATGAAGGATGAACATAGTGTGATTTCTAATCTCTCCCTT 2773
Db 697 GTTTGAAGCCAGCCCTGATAATGAAGGATGAACATAGTGTGATTTCTAATCTCTCCCTT 638

Qy 2774 TTTTGATTTAGTTGGATGTCTTTTAAATGTCTTTTCCCTGCATGAGTGGAAAGGGGAC 2833
Db 637 TTTTGATTTAGTTGGATGTCTTTTAAATGTCTTTTCCCTGCATGAGTGGAAAGGGGAC 578

Qy 2834 CTTTTTGAGTTGTCAATTTTGCACCTTTCAAACTTATTTTCTTGGAAACAAATATTTATAG 2893
Db 577 CTTTTTGAGTTGTCAATTTTGCACCTTTCAAACTTATTTTCTTGGAAACAAATATTTATAG 518
```



QY 2894 GGCCTAAAGCCCATTTTCAATCTTAATCTAAATTAATGATGTCCTATCTG 2942  
|||||  
Db 517 GGCCTAAAGCCCATTTTCAATCTTAATCTAAATTAATGATGTCCTATCTG 469  
|||||

## RESULT 11

US-10-033-528-45/c

; Sequence 45, Application US/10033528

; Publication No. US20020131971A1

; GENERAL INFORMATION:

; APPLICANT: King, Gordon E.

; APPLICANT: Meagher, Madeleine Joy

; APPLICANT: Xu, Jiangchun

; APPLICANT: Secrist, Heather

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY

; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER

; FILE REFERENCE: 210121.547C1

; CURRENT APPLICATION NUMBER: US/10/033,528

; CURRENT FILING DATE: 2001-12-26

; NUMBER OF SEQ ID NOS: 1896

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 45

; LENGTH: 706

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc feature

; LOCATION: 34, 35, 39

; OTHER INFORMATION: n = A,T,C or G

US-10-033-528-45

Query Match 6.8%; Score 199; DB 13; Length 706;

Best Local Similarity 94.8%; Pred. No. 3.4e-48;

Matches 217; Conservative 0; Mismatches 10; Indels 2; Gaps 1;

QY 2716 GGTTTGAAGCCAGCCCTGATATGAAGATGAACCTAGTGTGATTTCTAATCCT--CCCTT 2773  
|||||

Db 697 GTTTGAAGCCAGCCCTGATATGAAGATGAACCTAGTGTGATTTCTAATCCTCCTCCCTT 638  
|||||

QY 2774 TTTTGAATTTAGTTGGATGTCCTTTTAAATGCTTTTGCCTGCATGAGTGGAAAGGGGAC 2833  
|||||

Db 637 TTTTGAATTTAGTTGGATGTCCTTTTAAATGCTTTTGCCTGCATGAGTGGAAAGGGGAC 578  
|||||

QY 2834 CTTTTCGAGTTGTCATTTTGCACCTTTCAAAACTTATTTTCTTGAAAAACAATATTATAG 2893  
|||||

Db 577 CTTTTCGAGTTGTCATTTTGCACCTTTCAAAACTTATTTTCTTGAAAAACAATATTATAG 518  
|||||

QY 2894 GGCCTAAAGCCCATTTTCAATCTTAATCTAAATTAATGATGTCCTATCTG 2942  
|||||

Db 517 GGCCTAAAGCCCATTTTCAATCTTAATCTAAATTAATGATGTCCTATCTG 469  
|||||

## RESULT 12

US-10-099-926-45/c

; Sequence 45, Application US/10099926

; Publication No. US2003016064A1

; GENERAL INFORMATION:

; APPLICANT: King, Gordon E.

; APPLICANT: Meagher, Madeleine Joy

; APPLICANT: Xu, Jiangchun

; APPLICANT: Secrist, Heather

; APPLICANT: Jiang, Yugu

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY

; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER

; FILE REFERENCE: 210121.547C2

; CURRENT APPLICATION NUMBER: US/10/099,926

; CURRENT FILING DATE: 2002-03-17

; NUMBER OF SEQ ID NOS: 1982

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 45

; LENGTH: 706

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: 34, 35, 39  
; OTHER INFORMATION: n = A,T,C or G  
US-10-099-926-45

## Query Match

Best Local Similarity 94.8%; Score 199; DB 16; Length 706;

Matches 217; Conservative 0; Mismatches 10; Indels 2; Gaps 1;

QY 2716 GGTTTGAAGCCAGCCCTGATATGAAGATGAACCTAGTGTGATTTCTAATCCT--CCCTT 2773  
|||||

Db 697 GTTTGAAGCCAGCCCTGATATGAAGATGAACCTAGTGTGATTTCTAATCCTCCTCCCTT 638  
|||||

QY 2774 TTTTGAATTTAGTTGGATGTCCTTTTAAATGCTTTTGCCTGCATGAGTGGAAAGGGGAC 2833  
|||||

Db 637 TTTTGAATTTAGTTGGATGTCCTTTTAAATGCTTTTGCCTGCATGAGTGGAAAGGGGAC 578  
|||||

QY 2834 CTTTTCGAGTTGTCATTTTGCACCTTTCAAAACTTATTTTCTTGAAAAACAATATTATAG 2893  
|||||

Db 577 CTTTTCGAGTTGTCATTTTGCACCTTTCAAAACTTATTTTCTTGAAAAACAATATTATAG 518  
|||||

QY 2894 GGCCTAAAGCCCATTTTCAATCTTAATCTAAATTAATGATGTCCTATCTG 2942  
|||||

Db 517 GGCCTAAAGCCCATTTTCAATCTTAATCTAAATTAATGATGTCCTATCTG 469  
|||||

## RESULT 13

US-10-961-527-45/c

; Sequence 45, Application US/10961527

; Publication No. US20050147615A1

; GENERAL INFORMATION:

; APPLICANT: King, Gordon E.

; APPLICANT: Meagher, Madeleine Joy

; APPLICANT: Xu, Jiangchun

; APPLICANT: Secrist, Heather

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY

; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER

; FILE REFERENCE: 210121.547C4

; CURRENT APPLICATION NUMBER: US/10/961,527

; CURRENT FILING DATE: 2004-10-07

; PRIOR APPLICATION NUMBER: US 09/920,300

; PRIOR FILING DATE: 2001-07-31

; PRIOR APPLICATION NUMBER: US 60/302,051

; PRIOR FILING DATE: 2001-06-29

; PRIOR APPLICATION NUMBER: US 60/279,763

; PRIOR FILING DATE: 2001-03-28

; PRIOR APPLICATION NUMBER: US 60/223,283

; PRIOR FILING DATE: 2000-08-03

; NUMBER OF SEQ ID NOS: 1789

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 45

; LENGTH: 706

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc feature

; LOCATION: 34, 35, 39

; OTHER INFORMATION: n = A,T,C or G

US-10-961-527-45

## Query Match

Best Local Similarity 94.8%; Score 199; DB 22; Length 706;

Matches 217; Conservative 0; Mismatches 10; Indels 2; Gaps 1;

QY 2716 GGTTTGAAGCCAGCCCTGATATGAAGATGAACCTAGTGTGATTTCTAATCCT--CCCTT 2773  
|||||

Db 697 GTTTGAAGCCAGCCCTGATATGAAGATGAACCTAGTGTGATTTCTAATCCTCCTCCCTT 638  
|||||

QY 2774 TTTTGAATTTAGTTGGATGTCCTTTTAAATGCTTTTGCCTGCATGAGTGGAAAGGGGAC 2833  
|||||

Db 637 TTTTGAATTTAGTTGGATGTCCTTTTAAATGCTTTTGCCTGCATGAGTGGAAAGGGGAC 578  
|||||

Qy 2834 CTTTGTGAGTGTGCTTTTGCACCTTCAAACTTATTTTCTTGGAACAATATTTATAG 2893  
 Db |||||  
 577 CTTTGTGAGTGTGCTTTTGCACCTTCAAACTTATTTTCTTGGAACAATATTTATAG 518  
 Qy 2894 GGCTTAAGCCCATTTTCATTTCTAACTAAATATGTGTCCTATCTG 2942  
 Db |||||  
 517 GGCTTAAGCCCATTTTCATTTCTAACTAAATATGTGTCCTATCTG 469

RESULT 14  
 US-09-983-965-2096  
 ; Sequence 2096, Application US/09983965  
 ; Patent No. US20020137160A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Warren, Wesley C.  
 ; APPLICANT: Tao, Mengbing  
 ; APPLICANT: Byatt, John C.  
 ; APPLICANT: Mathialagan, Nagappan  
 ; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND  
 ; TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION  
 ; FILE REFERENCE: 37-21(10297)C  
 ; CURRENT APPLICATION NUMBER: US/09/983,965  
 ; CURRENT FILING DATE: 2001-10-26  
 ; PRIOR APPLICATION NUMBER: US 09/465,231  
 ; PRIOR FILING DATE: 1999-12-15  
 ; PRIOR APPLICATION NUMBER: US 60/113,678  
 ; PRIOR FILING DATE: 1998-12-17  
 ; NUMBER OF SEQ ID NOS: 5912  
 ; SEQ ID NO 2096  
 ; LENGTH: 399  
 ; TYPE: DNA  
 ; ORGANISM: Bos taurus  
 ; FEATURE:  
 ; OTHER INFORMATION: Clone ID: 32-LIB3057-012-Q1-K1-H11  
 US-09-983-965-2096

Query Match 6.8%; Score 198.6; DB 9; Length 399;  
 Best Local Similarity 83.8%; Pred. No. 3e-48;  
 Matches 249; Conservative 0; Mismatches 44; Indels 4; Gaps 2;  
 Qy 2323 AAAGTAACTTCTTGAAGAGTACTGAGTATATATAGTGTGATGAGGATCGAA 2382  
 Db |||||  
 102 ATAGGTAACCTTCTTGAAGAGTACTGAGTATATATAGTGTGATGAGGATCGAA 161  
 Qy 2383 AGGACCATGGAGAAATTTGCAAGGATGGATGAGGTTCCAGAAACGAATTCAGAAAC 2442  
 Db |||||  
 162 AGGACCATGGAGAAATTTGCAAGGATGGATGAGGTTCCAGAAACGAATTCAGAAAC 221  
 Qy 2443 AGAAGATGCTATTCGATATTCCTTGACATTTGAAACACAGAGAAAGATGTTTAAATAGA-- 2500  
 Db |||||  
 222 AGAAGATGCTATTCGATATTCCTTGACATTTGAGCAGACAGAGAAAGATGTTTAAATAGA 281  
 Qy 2501 -CTCAGGGAACATGCTTCAAGGACTTAATGTTCTCAAGCAATGTTGAGTTG-GCAGCC 2558  
 Db |||||  
 282 CTTCAAGCAACATGTTTCAAGGACTTAATGTTTTCAGCAATGTTTAAAGATGATTCGAC 341  
 Qy 2559 TGTAGTCTAGCTAGCATACACTACCTTACCTGAGAGGTTCTTTTAAACAA 2615  
 Db |||||  
 342 AGCCTCTGGCCCTAGCATACACTACCTTACCTGAGAGGTTCTTTTAAACAA 398

RESULT 15  
 US-09-833-790-371  
 ; Sequence 371, Application US/09833790  
 ; Patent No. US20020068288A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lodes, Michael J.  
 ; APPLICANT: Wang, Tongtong  
 ; APPLICANT: Secret, Heather  
 ; APPLICANT: Mohamath, Raodoh  
 ; APPLICANT: Indirias, Carol Y.  
 ; APPLICANT: Fan, Liqun  
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY

; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER  
 ; FILE REFERENCE: 210121-512  
 ; CURRENT APPLICATION NUMBER: US/09/833,790  
 ; CURRENT FILING DATE: 2001-04-11  
 ; NUMBER OF SEQ ID NOS: 440  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 371  
 ; LENGTH: 628  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; ORGANISM: Homo sapiens  
 US-09-833-790-371  
 Query Match 6.4%; Score 187.4; DB 9; Length 628;  
 Best Local Similarity 99.5%; Pred. No. 9.5e-45;  
 Matches 188; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 Qy 2754 GTGATTTCTAAATCCTCCCTTTTGTGATTTAGTTGGATGCTTTTAAATGTCCTTTGCCT 2813  
 Db |||||  
 1 GTGATTTCTAAATCCTCCCTTTTGTGATTTAGTTGGATGCTTTTAAATGTCCTTTGCCT 60  
 Qy 2814 GCATGAGGTGGAAGGGGACCTTTTGTGATTTGTCTTTTGCACCTTTCAAAACTTATTTTC 2873  
 Db |||||  
 61 GCTTGAGGTGGAAGGGGACCTTTTGTGATTTGTCTTTTGCACCTTTCAAAACTTATTTTC 120  
 Qy 2874 TTGGAACCAATATTTATAGGCTTAAAGCCCATTTTCAATTTCTAATCTAAATATATGTGT 2933  
 Db |||||  
 121 TTGGAACCAATATTTATAGGCTTAAAGCCCATTTTCAATTTCTAATCTAAATATATGTGT 180  
 Qy 2934 GCCTATCTG 2942  
 Db |||||  
 181 GCCTATCTG 189

Search completed: September 16, 2005, 10:39:57  
 Job time : 1746.36 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 15, 2005, 11:51:14 ; Search time 8782.62 Seconds  
(without alignments)  
12750.757 Million cell updates/sec

Title: US-10-650-482-1  
Perfect score: 2942  
Sequence: 1 attttggcttcgtccac.....aaattatgtgcttatctg 2942

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

- 1: gb\_est1.\*
- 2: gb\_est2.\*
- 3: gb\_hic.\*
- 4: gb\_est3.\*
- 5: gb\_est4.\*
- 6: gb\_est5.\*
- 7: gb\_est6.\*
- 8: gb\_gest1.\*
- 9: gb\_gest2.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1605.4	54.6	3941	3	HSM805059
2	1364.4	46.4	2619	3	AK082957 Mus muscu
3	1363.4	46.3	2731	3	AK049028 Mus muscu
4	919.4	31.3	979	1	AL526219 Mus muscu
5	868.8	29.5	897	5	BU162260 AGENCOURT
6	834.8	28.4	873	5	BQ221596 AGENCOURT
7	803	27.3	835	6	CD656897 AGENCOURT
8	777.2	26.4	1021	4	BM559080 AGENCOURT
9	768.2	26.1	775	6	CD642687 AGENCOURT
10	758	25.8	793	6	CA749002 UI-H-FE1-
11	755.6	25.7	795	6	CA426770 UI-H-FE1-
12	748.8	25.5	850	4	BG721970 602698720
13	736.4	25.0	875	5	BQ427697 AGENCOURT
14	734.6	25.0	838	7	CF594071 AGENCOURT
15	728.4	24.8	793	6	CB989883 AGENCOURT
16	719.4	24.5	792	7	CF538482 AGENCOURT
17	712.6	24.2	1047	2	BF797553 602257337
18	710	24.1	776	1	AI123720 ool16h12.x
19	709	24.1	721	5	BX510286 DfZp781H
20	707.6	24.1	830	6	CB960555 AGENCOURT
21	699.8	23.8	764	6	CB990340 AGENCOURT
22	692	23.5	786	4	BG434646 602506723
23	685.8	23.3	784	6	CB960059 AGENCOURT
24	678	23.0	1637	3	CR595060 full-leng

98	487.2	16.6	709	7	CF745488	CF745488 UI-M-GVO-
99	479.6	16.3	716	4	BI088423	BI088423 602852920
c 100	473.2	16.1	478	1	AI684846	AI684846 wa86c12.x
ALIGNMENTS						
RESULT 1						
HSM805059						
LOCUS	3941 bp mRNA linear HTC 22-SEP-2004					
DEFINITION	Homo sapiens mRNA; cDNA DKFZp6661186 (from clone DKFZp6661186).					
ACCESSION	AL833746					
VERSION	AL833746.1 GI:21734396					
KEYWORDS	HTC.					
SOURCE	Homo sapiens (human)					
ORGANISM	Homo sapiens					
REFERENCE	Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;					
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.					
	1 (bases 1 to 3941)					
	Bahr, A., Lauber, J., Mewes, H. W., Weil, B., Amid, C., Osanger, A.,					
	Fobo, G., Han, M. and Wiemann, S.					
CONSRMT	The German cDNA Consortium					
TITLE	Direct Submission					
JOURNAL	Submitted (22-SEP-2004) MIPS, Ingolstaedter Landstr.1, D-85764					
COMMENT	Neuberberg, GERMANY					
	Clone from S. Wiemann, Molecular Genome Analysis, German Cancer					
	Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;					
	sequenced by Qiagen (Hilden/Germany) within the cDNA sequencing					
	consortium of the German Genome Project.					
	This clone (DKFZp6661186) is available at the RZPD Deutsches					
	Reisourcenzentrum fuer Genomforschung GmbH in Berlin, Germany.					
	Please contact RZPD for ordering:					
	http://www.rzpd.de/cgi-bin/products/cl.cgi?cloneID=DKFZp6661186					
	Further information about the clone and the sequencing project is					
	available at http://mips.gsf.de/projects/cdna/.					
FEATURES	Location/Qualifiers					
source	1..3941					
	/organism="Homo sapiens"					
	/mol_type="mRNA"					
	/db_xref="RZPD:DKFZp6661186"					
	/db_xref="taxon:9606"					
	/clone="DKFZp6661186"					
	/tissue_type="stomach"					
	/clone_lib="666 (synonym: hsto2). Vector pSport1; host					
	DH10B; sites NotI + SalI"					
	/dev_stage="adult"					
	/note="hypothetical protein, N-terminus truncated"					
gene	1..3941					
	/gene="DKFZp6661186"					
CDS	42..1225					
	/gene="DKFZp6661186"					
	/codon_start=1					
	/product="hypothetical protein"					
	/protein_id="CAH56240.1"					
	/db_xref="GI:52545825"					
	/translation="ASKQQLPTPDQNGVHSLREHSLLRMDPKHCEDNPTQFVPA					
	GDIPNTQSTEEKIELTIVPLALEESPEGCPSEIPMEKEPGEGRISVVDISY					
	LEGDLPI SARPA CNKLIDYILGASDLETSSDPEGDWDEAEDDGFDSLSLDS					
	DLEQDPEGLHWNFCSDVPYNPNFTATIQTAAIRIVPEEPDSKDLGSKSLDENS					
	QSGSLPTEPHSSGEEDWESSADEASLKLWNSFCNSDDPNPLNFKAPFOTSGENE					
	KGKDSKTPSPSIVAISSCTLLSKVOLLGSEQCPDSVQRDVLGSGRRHTVKKRK					
	VTFLEEVTEYIISGEDRKWGFEPFARDGCRFQRQIETEDAIGYCLTFEHRERFNR					
	LQGTCKGLNVLKQC"					
ORIGIN						
Query Match	54.6%; Score 1605.4; DB 3; Length 3941;					
Best Local Similarity	99.9%; Pred. No. 0;					
Matches 1617; Conservative	0; Mismatches 1; Indels 1; Gaps 1;					
QY	1324 GGCTAACAGGGGCAAGATTATACCCACCCCTGAC CAGAGATAATGGCTACCA CAGCCTGGA 1383					
Db	1 GGCTAGCAAGGGGCAAGATTATACCCACCCCTGAC CAGAGATAATGGCTACCA CAGCCTGGA 60					

Qy	2464	CTTGACATTTTGAACACAGAGAAAGAAATGTTTAATAGACTCCAGGGAACATGCTTCAAGG	2523
Db	1141	CTTGACATTTTGAACACAGAGAAAGAAATGTTTAATAGACTCCAGGGAACATGCTTCAAGG	1200
Qy	2524	ACTTAAATGTTCTCAAGCAATGTTGAGTTGGCAGCTGTAGTCCCTAGCTAGCATACACTAC	2583
Db	1201	ACTTAAATGTTCTCAAGCAATGTTGAGTTGGCAGCTGTAGTCCCTAGCTAGCATACACTAC	1260
Qy	2584	CTCTTACCTGAGAGGTGTCTTTTAAACAACTTGGCAGCTGTCTTTGACATTTTTT	2643
Db	1261	CTCTTACCTGAGAGGTGTCTTTTAAACAACTTGGCAGCTGTCTTTGACATTTTTT	1319
Qy	2644	TTTTTACGGAATGTAACCTGGATCTAGTTTAAATTTTTTTTTTGGCAATATCCCACT	2703
Db	1320	TTTTTACGGAATGTAACCTGGATCTAGTTTAAATTTTTTTTTTGGCAATATCCCACT	1379
Qy	2704	CAGAAACATTCAGGTTTGAAGCCAGCCCTGATATGAAGATGAATAGTGTGATTTCTA	2763
Db	1380	CAGAAACATTCAGGTTTGAAGCCAGCCCTGATATGAAGATGAATAGTGTGATTTCTA	1439
Qy	2764	ATCCTCCCTTTTGTAGTTAGTTGGATGTGCTTTTAAATGTCTTCCCTGCATGAGGTG	2823
Db	1440	ATCCTCCCTTTTGTAGTTAGTTGGATGTGCTTTTAAATGTCTTCCCTGCATGAGGTG	1499
Qy	2824	GAAGGGGACCTTTTTCAGTTGTCATTTTGCACCTTTCAAACTTATTTCTTGGAAACA	2883
Db	1500	GAAGGGGACCTTTTTCAGTTGTCATTTTGCACCTTTCAAACTTATTTCTTGGAAACA	1559
Qy	2884	ATATTTATAGGGCTTAAAGCCCATTTTCATTTCTAACTAAATATGTGTGCTTATCTG	2942
Db	1560	ATATTTATAGGGCTTAAAGCCCATTTTCATTTCTAACTAAATATGTGTGCTTATCTG	1618
RESULT 2			
LOCUS	AK082957	2619 bp mRNA linear HTC 03-APR-2004	
DEFINITION	Mus musculus 12 days embryo spinal cord cDNA, RIKEN full-length enriched library, clone: C530022L24 product: hypothetical protein, full insert sequence.		
ACCESSION	AK082957		
VERSION	AK082957.1	GI:26350136	
KEYWORDS	HTC; CAP trapper.		
SOURCE	Mus musculus (house mouse)		
ORGANISM	Mus musculus		
REFERENCE	1 Carninci, P., and Hayashizaki, Y.		
AUTHORS	Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.		
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes		
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)		
MEDLINE	20499374		
PUBMED	10349636		
REFERENCE	2 Carninci, P., and Hayashizaki, Y.		
AUTHORS	Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M., Sugahara, Y., Shibata, K., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsunoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaki, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, M., Yoneda, Y., Ishikawa, T., Okawa, K., Tanaka, T., Matsunaga, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.		
TITLE	RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer		
JOURNAL	Genome Res. 10 (11), 1757-1771 (2000)		
MEDLINE	20530913		
PUBMED	11076861		

REFERENCE	
AUTHORS	4 The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
TITLE	Functional annotation of a full-length mouse cDNA collection
JOURNAL	Nature 409, 685-690 (2001)
REFERENCE	5 The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
AUTHORS	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
TITLE	Nature 420, 563-573 (2002)
JOURNAL	6 (bases 1 to 2619)
AUTHORS	Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, M., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, K., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akashira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M., and Hayashizaki, Y.
TITLE	Direct Submission
JOURNAL	Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/, tel: 81-45-503-9222, Fax: 81-45-503-9216)
COMMENT	cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
	Please visit our web site for further details.
	URL: http://genome.gsc.riken.jp/
	URL: http://fantom.gsc.riken.jp/
FEATURES	Location/Qualifiers
source	1..2619
	/organism="Mus musculus"
	/mol_type="mRNA"
	/strain="C57BL/6J"
	/db_xref="FANTOM DB:C530022L24"
	/db_xref="taxon:10090"
	/clone="C530022L24"
	/tissue_type="spinal cord"
	/clone_lib="RIKEN full-length enriched mouse cDNA library"
	/dev_stage="12 days embryo"
	418..2511
	/note="unnamed protein product; hypothetical protein (evidence: ProCrest)
	putative"
	/codon_start=1
	/protein_id="BAC38708.1"
	/db_xref="GI:26350137"
	/translation="METGTHRAKRPGRPLGWSFRLPFLRRSHACSSPPPPSRQNP
	GNALPERTRYWTKLLSQLALLPSLFLQKLLMLWSQLFGGLIPTRLDPAASYSALRA
	LKREESAPTVOKLSLSSLRSLDVLVSSLDLWEEGLQWQCSSDLKLAQRA
	LSAAPTFLLEQLQWGLVPSLQAGLVSHRELDSSSGPLSVQSLGNPKVYSYLLN
	PSYLDLPQLGRCSAGGQGVFGRTITPSCYLSDEGCHQHPQRAEMSAFWRRRC
	PLSTGLPEIHLRMKRLFLQANKQBLPTPDQNGVHLSLEENHLLRMDPHCTD
	NPAQVSPADRPETKEKPELVIOVSQSGSSLFCELPEVECEEDHTNATDLSQ
	RESLVPSTRPVCSNKLIDYLGAPSDLEASDSESDWGEPEPDGDSGLSES
	DVEQDSGLHWNPSFVDIPNPQNTATQTAARAPRDPDSVSGNTPKFPQSGCQ
	EGPLPTPHSSGEEDDWEPSDAENLKLWNSFCHSEDPYLLNLPKFPQSGKWK
	GRDQSKASSEATVAFSGHHTLLSCAKLLESDQNCPCGGLAGLAGEYTHIKKKV
	TFLEVTYVYISGDEORKGWEEFARDGCRFKRIQIETVAIGCYCLAFEHKQFNRL
	RIESKDLLLYSNVKK"
ORIGIN	

Query Match		46.4%;	Score 1364.4;	DB 3;	Length 2619;
Best Local Similarity		73.9%;	Pred. No. 0;		
Matches 1900;		Conservative	0;	Mismatches 611;	Indels 59; Gaps 11;
Qy	96	CCATGCGCGAGCCGTGGGGGACCTCTATCCATGGTGTGAAGCGTCGAGCCGACATAGG	155		
Db	107	CGCTCGCGCGCGAGGAGGGGTCTTCTATATGGTGAGCGATCTCACACGCGCTAGG	166		
Qy	156	GAACCTCTCTCCC-CGCCAGGATGGAAGTCAGTCAGTCGCGGCTATGCGCGGCTGT	214		
Db	167	ACGTCTCTCTCCCTAGCGGATGGACCTAACCGCGCTCGCACCGCTTGCGCGGCTC	226		
Qy	215	TCCTCCCTGTGTCTGCGCGCGCTGCGCATTTGCTGCCCTCTGTGTGCTTTCTGTGG	274		
Db	227	TGGGCGCTCCGGTCAGCACTGTGTGCGAAGCGCGCTCTCTGGGCTCTCTGCGG	286		
Qy	275	CTCGAAGATCGGCTTGGAGCGACGACCGCCGCTGGGCAAGCGCGAGACTCTGTAGGCT	334		
Db	287	CGCGGAATCGACTGCAGTACCACTCCGTGGCTGGCAAGCGGAGACTGTGTAGACC	346		
Qy	335	TCCTCCGAATCCGTCGACCTCCAGCCGCTGAGCGCGCGGCGCTACTCTGAGAGACTGTC	394		
Db	347	TCGATCCAGCTCGCTGACCGCGCTGAGCTCTGCTCTCTCTGTGTGAGAAGCGGCC	406		
Qy	395	AAGAAAAGGAGATGGAGCCGGGACAGCGGATCGCGAAAACGGCTTTGGCCCTCGGGCG	454		
Db	407	AAG-GAAAGGAGATGGAGACAGGAACGACAGGCGCCGGAAGCGGCTTGGCCCTCGGCTG	465		
Qy	455	GGCTTCGGTTCGGCCACCTTTTTCCTCGGGATCGCAAGCAGGCTCTCTTAAGTTC	514		
Db	466	GGCTCTCGTTCGGCTGCC- --TTCCTTGGGGATCGCACGCGCTGCTCTTCGGAGTTC	522		
Qy	515	CAGAGCTCTTGGCCCGGAAATCCCGGGAACCCACACTGCTTCTCTGCGCCAGGCC	574		
Db	523	CCGCGGCTCTCTCTCGACNAATTCGCGGAA- -----CTCGCTCTGCCC	567		
Qy	575	GAGACTCGGTCAGTTACTGGACGAACTGCTCTCCAGCTCTTTCGCGCCCTCCCGGA	634		
Db	568	GAGGTCGACACGATCTGACCCAAATGCTTCTAGCTCTTTCGCTCTCCCTAGC	627		
Qy	635	TGCTTCAGAAAGGTGCTAATTTGGAGCCAACTTTTCGGTGGAAATGTTTCGACAGATGG	694		
Db	628	CTATTCCAGAAAGCTGCTCTTTGGAGCAGCTTTTCGGGGCGCTGATTTCTTACCAGATGG	687		
Qy	695	CTAGATTTGCTGGAGTCTACAGCGCCTGAGAGCCCTGGAAGGACGGGAAACACCGCC	754		
Db	688	CTAGATTTGCGCAAGTTACAGGCGCCTGAGAGCTTTTGAGAGGACGGGAAATCTGCC	747		
Qy	755	GCCCCACAGCGCAGAAATCTTTGAGTTGCTGCGAGCTCGACTCTCTCAGACCCCTCGGTC	814		
Db	748	GCTCCACGCTGACAGACTCTCTGAGTTCACTGCGGCTGGACTCTTCGGAAGACTTGGTT	807		
Qy	815	ACGAGTCCCTTGATTTGGCTAGAGAGGGGATCCAATGGCAATATCTGCCGCCACGACTTA	874		
Db	808	GTCAGTTCTCTTGATTTGGCTAGAGAGGGAATCCAGTGGCAGTGTCTCTCAGACCTG	867		
Qy	875	AAATTTGAGCTTAAGGCCAAGGAGTCTTTGACCTCTGACGACAGCGCTTTCTCTTA	934		
Db	868	GAGTTAAACATCAAGGCCCAGAAAGACTTTAGACTCTGACGCGCCCACTTTCTCTCG	927		
Qy	935	GAGCAGAGCTGTGGGAGTGGAGCTGTTGGCCAGTAGCTTCAATCCCGCTCTGTACTCT	994		
Db	928	GAGCAGAGCTGTGGGAGTGGAGTTGCTGCCCAGTAGCTTCAAGCTGGTCTAGTCTCC	987		
Qy	995	AACGGGAACCTTGGCTCTTGGCCCTCTGGGCTCTAAACATTTCAACGATAGACAAATTC	1054		
Db	988	CACCGAAGATTTGACTCTTCTCATCTCTGGGCTCTGAGCGTTTCAGAGCTTAGGTAATTC	1047		
Qy	1055	AGTGTGTATCCTATTTGCTGAAACCTTCTTACCTGAGCTGCTTTCTTAGGCTAGAGTTC	1114		
Db	1048	AAGGTAGTTTCTCTATCTCTGAAACCTTCTTACCTGGACTACCTTCCCGAGTTAGGGCTG	1107		
Qy	1115	AGCTATCAGAAACAGTGATGGAATAGCGAGGTAGTCGGCTTTCAGACACTAACCCACAG	1174		

Db	1108	CGCTGTCCAGAGCAGCGCTGGAGGTGGCAGTTTGTGGTTTCCGAACTAACCCACAG	1167		
Qy	1175	AGCAGCTGCTGAGAGGAGACCATTTGTTCATCCCAGCGCTGAGTGAGAACTCATTTCCG	1234		
Db	1168	AGCTGCTATCTTTCTGAAGATGGTTGTACACCTCAGCCGTTCGGGGCAGAGATGTCGGCA	1227		
Qy	1235	GCCTCGTGGCAGGATGTCACACTCTTTCTACGGAAGCCCTACCAGAAATTCACCATCTT	1294		
Db	1228	ACGCTGAGAGAGGTGTCGCTCTCTCTACAGAGGCTTGCAGGAAATCCACCATCTT	1287		
Qy	1295	CGCATGAAAACGGCTGGAAATTCCTTCAACAGGCTAAAGGGGCAAGATTTACCCACCCCT	1354		
Db	1288	CGTATGAAAACGGCTAGAAATTCCT- --CCAGGCTAAACAAAGGCAAGAGTTACCCACCCCT	1344		
Qy	1355	GACGAGATTAATGGCTACACAGCTGGAGGAGGACACAGCCTTCTCCGATGGATGCCA	1414		
Db	1345	GACCAAGATTAATGGCTATATAGCTTGGAGGAGGAAATTAACCTTCTCCGATGGAGCCCA	1404		
Qy	1415	AAACACTGACAGAGATAACCAACACAGTTTGTCTCTGCTGAGAGACATTTCTTGGAAAC	1474		
Db	1405	CAACATTTGACAGATAACCCAGCACAGGCGGTGTCCTGCTGCAGACAGAGCC- -----	1457		
Qy	1475	ACCCAGGAATCCAATGAAGAAAATAAGAAATTAATACTACAGAGGTTCACATTTGCTTTG	1534		
Db	1458	-----GGAGCCCACTGAGAAAAAAACACAGAAATTTGGTGAATCAAGAAAGTTTCAC	1504		
Qy	1535	GAAGAGAGAGCCCTTCTGAGGCTGTCCATCTAGTGAGATACCTATAGAAAAGAGGCGCT	1594		
Db	1505	-----AGAGCCCCCAGGGAAGCAGTCTGTTTGTGAATTAACCCGTGGAAAAAGAAATGT	1557		
Qy	1595	GGAGAGGCGCAATAAGTGTAGTTGATTACTCATACCTAGAAAGTGACCTTCCCATTTCT	1654		
Db	1558	GAAGAGAACACACTAATGCAACTGACCTCTCAGATAGAGAGAGAGCCCTTCTGTTTCT	1617		
Qy	1655	GCCAGACAGCTTTGTAGTAAACAACTGATAGATTAATAATTTGGAGGTGCAATCCAGTGAC	1714		
Db	1618	ACCAGACAGTTTGTAGCAACAACTGATAGATTAATAATTTGGAGGCGCCCCAGTGAC	1677		
Qy	1715	CTGGMAACAAGTTCGATCCAGAGGTGAGGATGGGATGAGGAAGCTGAGGATGATGGT	1774		
Db	1678	TTGGAGCCAGCTCTGATTTCTGAAAGTGAGGATTTGGGGCGAGGAACCTGAGGACGATGGC	1737		
Qy	1775	TTTGATAGTATAGCTCACCTGTGAGACTCAGACCTTGAAACAAAGACCCCTGAAGGCTTCAC	1834		
Db	1738	TTTGATAGCGATGGCTCCCTGTCTGATCAGACGTGGAACAGGACTCGGAAGGCTTTCAC	1797		
Qy	1835	CTTTGGAACTCTTTCTGCAGTGATGATCTTTATAATCCCGACAGAACTTTACAGCAAAAT	1894		
Db	1798	CTTTGGAACTCTTTCCACAGTGTAGATCTTTACAAACCCCAAACTTTTACAGCCACGATT	1857		
Qy	1895	CAGACTGCTGCCAGAAATTTCTTGAGAGCCCTTCTGATTCAGAGAGGATTTGCTCTGGC	1954		
Db	1858	CAGACGCTGCCAGAAATTTGCCCGACAGACCCATCAGATTCAGGGAATCTCTGCTCTGGC	1917		
Qy	1955	AAGTCTGATCTAGAGAAATTCCTCCAGTCTGGAAGCCCTTCTGAGACCCCTGAGCATAGT	2014		
Db	1918	AGCTGTGTGTAGGA- --CTGTGAGAGGAGACCCCTTCCGGAGACCCCGACCATAGT	1974		
Qy	2015	TTTGGGAGGAAGATGATCTGGGAATCTAGTGAGATGAAGCAGAGAGTCTCAAACTGTGG	2074		
Db	1975	TCCGGGAGGAAGATGATCTGGGAACCGAGTGCAGATGAAGCAGAGAAATCTTAAATTTGG	2034		
Qy	2075	AACCTATTCTGTAATTTCTGATGACCCCTCAACCTTTAAATTTTAAGGCTCTTTTCAA	2134		
Db	2035	AACCTTTTCTGTCAATCTGAGGACCCCTCAACCTTTAAATTTTAAGGCTCTTTTCAA	2094		
Qy	2135	ACATCAGGGGAAATGAGAAAGGCTGTCTGACTCAAGAGCCCCCATCTGAGTCCATTTGTG	2194		
Db	2095	CCGTACGGGAAGAAATTGGAAGGCGGTGAGACTCAAGAGCCCTCTCTGAGGCTACAGTG	2154		
Qy	2195	GCCATTTCTGAGTGTACACCTTACTTTCTTGTAGGTGAGCTGTTGGGAGGCAAGAA	2254		



2155	Db	GCCTTCTCTGGCCATCATACCTTACTTCTTTGTAAGGCCCCAGCTGTTAGAGAGCCAGAA	2216
2255	QY	AGTGAATGTCACGACTCGGTACACGCGTGACGTTCTTTCTGGAGGAAGACACACACATGTC	2314
2215	Db	GATAATTGTCACGGCTGTGGCTGGGTGAGGCTCTTCTGAGAGAAAGATACACCCATATC	2274
2315	QY	AAAAGAAAAAGGTAACTTCTCTTGAAGAAGTACTGAGTATTATATAAGTGGTGATGAG	2374
2275	Db	AAGAGAAAAAGGTAACTTCTCTGGAAGAAGTACTGAGTATTATATAAGTGGTGATGAG	2334
2375	QY	GATCGCAAGAGGCATCGGGAAGAATTTGCAAGGGATGCGATGCGAGGTTCCAGAAACGAATT	2434
2335	Db	GATCGCAAGAGGCATCGGGAAGAATTTGCAAGGGATGCGATGCGAGGTTCCAGAAACGAATT	2394
2435	QY	CAAGAAACAGAAAGTGTATTGGATATTTGCTTGACATTTTGAACACAGAGAAAGATGTTT	2494
2395	Db	CAAGAAACAGAAAGTGTCCATTGGCTACTGCTTGGCCCTTTCGAGCACAGAGAAAAATGTTT	2454
2495	QY	AATAGACTCAGGGAACATGCTTCAAAGGACTTAATGTTCTCAAGCAATGTTGAGTTG-G	2553
2455	Db	AATAGACT----GAGGATCGAGTCAAAGGACTTACTGTTGTACAGCAATGTTAAGAAGTG	2510
2554	QY	CAGCCTGTAGTCTCTAGCTAGCATACACTCTTCTTACCTCAGAGGCTGCTCTTTTAAAAACA	2613
2511	Db	AACAGCCCTCAAAGCCCGTCCCACTCTGCTCTTACTTGAGA-GTTTCCTTAAAAACA	2569
2614	QY	AATCTTGGCAGCTGCTCTTTTGACATTTTTTTTTTTTAGAGGAAATGTAACT	2663
2570	Db	AACACTGGCAGCTGCTCTTGGACATGTTTTTAAAGAAACAACCTTGTATCT	2619
RESULT 3			
AK049028			
LOCUS			
DEFINITION			
AK049028 2731 bp mRNA linear HTC 03-APR-2004			
Mus musculus 0 day neonate cerebellum cDNA, RIKEN full-length			
enriched library, clone:C230093D15 product:hypothetical protein,			
full insert sequence.			
ACCESSION			
AK049028.1 GI:26339691			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTH			

Matches 1957; Conservative 0; Mismatches 641; Indels 79; Gaps 13;

Qy	123	TATCATGTGTGAAGCGTGCAGCGGCTAGGAACTCTCTTCC-CCGAGAGTGA	181
Db	2	TCTCTATGTGAGGATCTCACGGCTAGGAGCTCTCTCTTCCCTAGCGGATGGAC	61
Qy	182	GTCCGATCAGTCGCGCCCTATTGCGGGCTGTCTTCCCTGTGTCTGCGCGCGTGC	241
Db	62	CTAACCGCGTGCACCGCTTGGCGGGCTCTGGCGCTCGCGTGCAGACCTGTTGC	121
Qy	242	CGCATTCGTCGCCCTCTGTGGCTTTCTGTGCTCGAAGATCGGCCCTGGAGCAGCAG	301
Db	122	GGAAAGCCGCTCTCTGCGCTCTCTGCGGCTCTCTGCGGCGGGAAATCGGACTGCACTACCCACT	181
Qy	302	CCACCGCTGGGCAAGCGCGAGACTCTGTAGGCTTCTCCGAAATCCGCTCGACCTCCAGCC	361
Db	182	CGTGCTGGGCAAGCGGAGACTGTGTAGACCTCGGATCCGATCCGCTGCGCTGACGCCGCT	241
Qy	362	GCTGAGCGCGCGGCTCTACCTGAGAGACTGTCAAGAAAGGAGATGGAGCCGGGACA	421
Db	242	GAGCTCTGTCTCTCTCTGTCTGAGAGCGCCCAAG-GAAGGAGATGGAGCAGGACG	300
Qy	422	GGCGATCGCGAAACGGCTTGGCCCTCGGGGGCTTCCGGTTCGGCCACCTTTTTC	481
Db	301	CACAGGGCCCGAAGCGGCTCGGCCCTCGGGCTCTGTGTTCCGGCTGCC---TTC	357
Qy	482	CCTCGCGATCGCAAGCGGCTCTTAAGTTCGCGAGCTCTTGGCCCGGAAACTCC	541
Db	358	CTTGGCGATCGACGCTCTCTTGGAGTTCGCGGCTTCTCTCGACAAATCC	417
Qy	542	GGGAACCCACACTGCTCTCTGCCAGCCCGAGACTCGGGTCAGTTACTGACGAAA	601
Db	418	GGGAACCTCGCTCT-----GCCGAGCTGGACACGAGTACTGGACCAA	462
Qy	602	CTGCTCTCCAGCTCTTGGCGCGCTCCCGGATTTGCTTCAAGAGTCTTAATTGGAGC	661
Db	463	TTGCTTTCTCAGCTCTTGGCCCTGCTCTAGCTATTTCCAGAACTGCTGCTTTGAGC	522
Qy	662	CAACTTTTCGGTGGAACTGTTCCGACAGATGGCTAGATTTTCTGAGTCTACAGCGCC	721
Db	523	CAGCTTTTCGGGCGCTGANTCTTACCAGATGGCTAGATTTTGGCGAAGTTACAGCGCC	582
Qy	722	CTGAGAGCCCTGAAGGGAGGGAGAAACAGCGCCGCCACAGCGCGAATACTTTTGAGT	781
Db	583	CTGAGAGCTTTGAGAGGAGCGGAGAACTCGCCGCTCCCAAGTGCAGAACTCTGAGT	642
Qy	782	TGCTGAGCTGAGCTCTCAGACCCCTCGGTACAGTCCCTTGTATGTTGCTAGAGAG	841
Db	643	TCACTGCGGCTGAGCTTTGCGAAGACTTGGTTGTCACTTCTTTGATTGGCTAGAGAG	702
Qy	842	GGGATCCACTGGCAATACCTCGCCCCAGACTTAAATTTGGAGCTTAAGGCCAAGGGAGT	901
Db	703	GAACTGCAATGGAGTCTGCTCTCAGACTGGAGTTTAAACTCAAGGCCCAAGAAAGA	762
Qy	902	GCTTTGGACCTTGACGACAGGCTTTCTTTAGAGCAGAGCTGTGGGAGTGGAGCTG	961
Db	763	GCTTTAGACTCTGAGCGCCACTTTCTCTGGAGCAGAGCTGTGGGAGTGGAGTTG	822
Qy	962	TTGCCAGTAGCTTCAATCCGCTGTGACTCTTAACCGGGAACTTGGCTCTTCCGCTCT	1021
Db	823	CTGCCAGTAGCTTCAAGCTGGTCTAGCTCTCCACCCAGAACTTGAATCTTCACTCTCT	882
Qy	1022	GGGCTCTTAAACATTTCAAGGATAGACAAATTTCAAGTGGTATCTTATTTCTCAACCT	1081
Db	883	GGGCTCTGAGGCTTCAAGGCTTAGGTAAATTTCAAGGTAGTTTCTTATCTCTCAACCT	942
Qy	1082	TCCTACCTGAGCTGCTTTCTTAGGCTAGAAGTCAAGCTATCAGAACAGTATGGAATAGC	1141
Db	943	TCCTACCTGAGCTACTTTCCCGAGTTAGGCTGCGCTGTCAAGCAGCGCTGGAGTGGC	1002
Qy	1142	GAGTAGTCGGTTCAGACACTAACCCAGAGCAGCTGCCCTGAGAGAGGACCATGT	1201
Db	1003	CAGTTTGTGGTTCGGAACACTAACCCAGAGAGCTGCTATCTTTCTGAAGATGGTTGT	1062

Qy	1202	CATCCCGCGCTGAGTCGAGAACTCATTTCCGGCTCTGTGGCAGGATGTCCACCTCTT	1261
Db	1063	CACCTCAGCGTTGGCGCAGAGATGTGGCAACCGCTGGAGAAAGTGTCCGCTCTC	1122
Qy	1262	TCTACGGAAGCCCTACAGAAATTCACATCTTTCGATGAAACGGCTGGAAATCTTTCAA	1321
Db	1123	TCTACGAAGGCTTCCGGAATTCACCACTTCTGATGAAACGGCTAGAAATCTCT---C	1179
Qy	1322	CAGGCTAACAGGGGCAAGATTTACCCACCCCTGACAGGATTAATGCTACCAAGCTG	1381
Db	1180	CAGGCTAACAAAGGGCAAGATTTACCCACCCCTGACCAAGATTAATGCTATCATAGCTG	1239
Qy	1382	GAGGAGAAACACAGCTTCTCCGATGGATCCAAACACATGTCAGAGATAACCCCAACAG	1441
Db	1240	GAGGAGAAACATTAACCTTCTCCGATGGACCCACATTTGCAAGATAACCCAGCAG	1299
Qy	1442	TTTGTCTCTGCTGCTGAGACATTTCTTGGAAACCCAGAAATCCACTGAAGAAATA	1501
Db	1300	CGGCTGCTCTGCTGACAGAGCC-----GGAGCCACTGAGAAAAAACCG	1347
Qy	1502	GAATTAATACTACAGAGTTCCACTTGTCTTGGAGAGAGAGCCCTCTGAGGGCTGT	1561
Db	1348	GAATTTGTGATTCAGAAAGTTTCAC-----AGAGCCCCAGGGAAGCAGT	1392
Qy	1562	CCATCTAGTCAGATACCTATGGAAAGGAGCCCTGGAGGGCCGAAATAAGTGTAGTTGAT	1621
Db	1393	CTGTTTGTGATTAACCGTGGAAAGAAATGTGAAGAGACCACTAATGCNACTGAC	1452
Qy	1622	TACTCATACCTAGAAGTGACCTTCCATTTCTGCCAGACCACTGTGTAGTAACAACTG	1681
Db	1453	CTCTCAGATAGAGAGAGAGCTTCTGTTTCTACAGACCACTTGTAGCAACAACTG	1512
Qy	1682	ATAGATTAATTTTGGAGGTGCATCCAGTGACTGGAAACAAATTTGTATCCAGAGGT	1741
Db	1513	ATAGATTAATTTTGGAGGCGCCCGCAGTGTGGAAGCAGCTCTGATTTCTGAAGT	1572
Qy	1742	GAGGATTTGGATGAGGAAGCTGAGGATGATGTTTGTATAGTATAGCTCACTGTCTAGAC	1801
Db	1573	GAGGATTTGGGCGAGGAACTGAGGACATGGCTTTGTATAGCATGGCTCCCTGTCTGAA	1632
Qy	1802	TCAGACCTTGAAACAGACCTGAAAGGCTTCACTTTTGGAACTCTTTCTGAGTGTAGAT	1861
Db	1633	TCAGACGTGAAACAGGACTCGGAAGGCTTCACTTGTGAACTCTTTTCCACAGTGTAGAT	1692
Qy	1862	CCTTAATATCCCGAACTTTACAGCAAAATTCAGACTGCTGCCAGAAATTTGTCTCGAA	1921
Db	1693	CCTTACAAACCCCAAACTTTACAGCCACGATTCAGACGCTGCCAGAAATTTGCCCCAGA	1752
Qy	1922	GAGCTTCTGATTCAGAGAAGGATTTGCTTGGCAAGTCTGATCTAGAGAATTTCTCCAG	1981
Db	1753	GACCATCAGATTCAGGGACATCTGCTGTGGCAGCTGTGTTGTAGGGA---GCTGTGAG	1809
Qy	1982	TCTGGAAGCTTCTGAGACCCCTGAGCAATGTTCTGGGAGGAAGATGATCGGAATCT	2041
Db	1810	GAGGGACCTTTCGGGAGACCCCGACCAATGTTTCGGGGAGGAAGATGATCGGAACCG	1869
Qy	2042	AGTCAGATGAAGCAGAGAGTCTCAAACTGTGAACTCATTTCTGTAATTTCTGATGACCC	2101
Db	1870	AGTCAGATGAAGCAGAGATCTTAAATTTGTGAACTCTTTCTGATCTGAGGACCC	1929
Qy	2102	TACAACTTTTAAATTTTAAAGGCTCTTTTCAAACTCAGGGGAAATGAGAAAGGCTGT	2161
Db	1930	TACAACTTTTAAATTTTAAAGGCTCTTTTCAACCGTCAGGGAAAGAAATGAGAAAGCGT	1989
Qy	2162	CGTGACTCAAAGACCCCATCTGATCCATTTGTGGCCATTTCTGAGTGTCACTTACTTT	2221
Db	1990	CAGGACTCAAAGGCTCTTCTGAGGCTACAGTGGCTTCTCTGCCCCTCATATACCTTACTT	2049
Qy	2222	TCTTGTAAAGTGCAGCTGTGGGGAGCCAGAAAGTGAATGTCCAGACTCGGTACAGCT	2281
Db	2050	TCTTGTAAAGGCCAGCTGTTAGAGAGCCAAAGAAAGATAATTTGTCCAGGCTGTGGGCTGGT	2109



Db 908 TGGAGCTGTGGCCAGTAGCCTTCAATCMTCTCTGTCTCTTAACWGGGAMTTGGCTCTY 967  
QY 1014 CGCCCTCTGGGC 1025  
Db 968 GCCCTCTGGGC 979

RESULT 5  
LOCUS BUI62260  
DEFINITION AGENCOURT\_7939455 NIH\_MGC\_67 Homo sapiens cDNA clone IMAGE:6172811  
5', mRNA sequence.  
ACCESSION BUI62260  
VERSION BUI62260.1 GI:22676170  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 897)  
AUTHORS NIH-MGC <http://mgi.nci.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
Tissue Procurement: ATCC  
cDNA Library Preparation: Life Technologies, Inc.  
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone Distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LHAM13543 row: f column: 12  
High quality sequence stop: 673.  
Location/Qualifiers  
1..897  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6172811"  
/tissue\_type="retinoblastoma"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH MGC\_67"  
/note="Organ: eye; Vector: pCMV-SPORT6; Site 1: NotI;  
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 1.75 kb. Library constructed by Life  
Technologies."

ORIGIN  
Query Match 29.5%; Score 868.8; DB 5; Length 897;  
Best Local Similarity 99.1%; Pred. No. 2.9e-233;  
Matches 884; Conservative 0; Mismatches 7; Indels 1; Gaps 1;  
QY 21 CGCACCGCGGCTACCCAGTCTTCGGGTATCGCGTTCAGGGGCTTTCAACCT 80  
Db 6 CGTCGAGCGGCTTACCCAGTCTTCGGGTATCGCGTTCAGGGGCTTTCAACCT 65  
QY 81 CTGTCACTCGGAACCATCCGCGAGCGCTGGGGGACTCTTCCATGTTGTGAAGC 140  
Db 66 CTGTCACTCGGAACCATCCGCGAGCGCTGGGGGACTCTTCCATGTTGTGAAGC 125  
QY 141 GTCGAGCGGCTAGGGAACCTCTTCCCGCCAGGATGGAAGTCGATCAGTCGCGGCT 200  
Db 126 GTCGAGCGGCTAGGGAACCTCTTCCCGCCAGGATGGAAGTCGATCAGTCGCGGCT 185  
QY 201 ATTGGCGGGCTGTCTTCCTGTGTCTTCGCGCGCGCTGCGGATTCGCTGCTCTGT 260  
Db 186 ATTGGCGGGCTGTCTTCCTGTGTCTTCGCGCGCGCTGCGGATTCGCTGCTCTGT 245  
QY 261 GCCTTTCTGTGGCTCGAAGATCGGCTCGAGCAGCAGCCACCGCTGGGCAAGGCG 320  
Db 246 GCCTTTCTGTGGCTCGAAGATCGGCTCGAGCAGCAGCCACCGCTGGGCAAGGCG 305

QY 321 AGACTCTGTAGGCTTCTCCGAATCCGTCGACCTCCAGCCGCTGAGCGCGCGCCCTA 380  
Db 306 AGACTCTGTAGGCTTCTCCGAATCCGTCGACCTCCAGCCGCTGAGCGCGCGCCCTA 365  
QY 381 CCTGAGAGACTGTCAAGAAAAAGGAGATGAGCGCGGAGACAGCGGATTCGGGAAACGGC 440  
Db 366 CCTGAGAGACTGTCAAGAAAAAGGAGATGAGCGCGGAGACAGCGGATTCGGGAAACGGC 425  
QY 441 TTGGCCCTCGGGGGGCTCCGGTTCTGGCCACCTTTTCCCTCGCGGATTCGCAAGCAG 500  
Db 426 TTGGCCCTCGGGGGGCTTCGGGTCTGGCCACCTTTTCCCTCGCGGATTCGCAAGCAG 485  
QY 501 GCTCTTCTAAGTTTCCCGACGCTCTTGGCCCGGAAAACTCCGGGAAACCCACACTGCTTT 560  
Db 486 GCTCTTCTAAGTTTCCCGACGCTCTTGGCCCGGAAAACTCCGGGAAACCCACACTGCTTT 545  
QY 561 CTTTGCACCGCGGAGACTCGGGTCACTTACTGAGCAAACTGCTCTCCAGCTCTCTTG 620  
Db 546 CTTTGCACCGCGGAGACTCGGGTCACTTACTGAGCAAACTGCTCTCCAGCTCTCTTG 605  
QY 621 CGCGCTCCCGGATTCGTTTCAAGAGGTGCTTAATTTGGAGCCAACTTTTGGTGAATGT 680  
Db 606 CGCGCTCCCGGATTCGTTTCAAGAGGTGCTTAATTTGGAGCCAACTTTTGGTGAATGT 665  
QY 681 TTCGACCAAGTGTAGATTTTCTGAGTCTACAGCGCCCTGAGAGCCCTGAAGGAC 740  
Db 666 TTCGACCAAGTGTAGATTTTCTGAGTCTACAGCGCCCTGAGAGCCCTGAAGGAC 725  
QY 741 GGAGAAACAGCGCGGCCCCCAGAGCGCAAACTTTTGGTGAATGTGCTGAGCTGAGTCTCT 800  
Db 726 GGAGAAACAGCGCGGCCCCCAGAGCGCAAACTTTTGGTGAATGTGCTGAGCTGAGTCTCT 785  
QY 801 CAGACCCCTCGGTCAACAGTCCCTTGAATTTGGCTAGAGGAGGGATCCACTGGCAATACT 860  
Db 786 CAGACCCCTCGGTCAACAGTCCCTTGAATTTGGCTAGAGGAGGGATCCACTGGCAATACT 845  
QY 861 CGCCCCCAGACCTAAATTTGGAGCTTAAGCCCAAGGAAAGTGTCTTTGGACC 911  
Db 846 CGCCCCCAGACCTAAATTTGGAGCTTAAGCCCAAGGAAAGTGTCTTTGGACC 897

RESULT 6  
LOCUS BQ221596  
DEFINITION AGENCOURT\_7569589 NIH\_MGC\_92 Homo sapiens cDNA clone IMAGE:6043767  
5', mRNA sequence.  
ACCESSION BQ221596  
VERSION BQ221596.1 GI:20402996  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 873)  
AUTHORS NIH-MGC <http://mgi.nci.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
Tissue Procurement: ATCC  
cDNA Library Preparation: Life Technologies, Inc.  
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone Distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LHAM13285 row: e column: 16  
High quality sequence stop: 653.  
Location/Qualifiers  
1..873  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"

FEATURES  
source

ORIGIN		Query Match		Best Local Similarity		Matches		Conservative		Indels		Gaps		2;	
		28.4%; Score 834.8; DB 5; Length 873;		99.5%; Pred. No. 1.2e-223;		0; Mismatches		2;							
<div> <div> <div>/clone="IMAGE:6043767"</div> <div>/tissue_type="embryonal carcinoma, cell line"</div> <div>/lab_host="DH10B (phage-resistant)"</div> <div>/clone_lib="NIH_MGC 92"</div> <div>/notes="Organ: testis; Vector: pCMV-SPORT6; Site 1: NotI; Site 2: SalI; Cloned unidirectionally; oligo-AT primed. Average insert size 2.5 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH_MGC Library."</div> </div> </div>															
QY	1488	CTGAAGAAAAATAGAAATTATTAACCTACAGAGTTCCACTTGTCTTTTGGAAAGAGAGAGCC	1547												
DB	1	CTGAGAAAAAATAGAAATTATTAACCTACAGAGTTCCACTTGTCTTTTGGAAAGAGAGAGCC	60												
QY	1548	CTTCTGAGGCTCTCCATCTAGTGAGATACCTATGGAAGAGGAGCTGGAGAGGGCCGAA	1607												
DB	61	CTTCTGAGGCTCTCCATCTAGTGAGATACCTATGGAAGAGGAGCTGGAGAGGGCCGAA	120												
QY	1608	TAACTGTAGTTGATTACTCATCTAGTAAGGTGACCTTCCCATTTCTGCCAGACCCAGCTT	1667												
DB	121	TAACTGTAGTTGATTACTCATCTAGTAAGGTGACCTTCCCATTTCTGCCAGACCCAGCTT	180												
QY	1668	GTAGTTAAACAACTGATAGATTATTTTGGAGGTGCATCCAGTGACCTCGAAACCAAGTT	1727												
DB	181	GTAGTTAAACAACTGATAGATTATTTTGGAGGTGCATCCAGTGACCTCGAAACCAAGTT	240												
QY	1728	CTGATCCGAAGGTGAGGATTTGGATGAGGAAGCTGAGAGTATGTTTGTATGATGATA	1787												
DB	241	CTGATCCGAAGGTGAGGATTTGGATGAGGAAGCTGAGAGTATGTTTGTATGATGATA	300												
QY	1788	GCTCACTGTGAGACTGAGACTTTGAACAGACCTGAGGGCTTCACTTTTGGAACTCTT	1847												
DB	301	GCTCACTGTGAGACTGAGACTTTGAACAGACCTGAGGGCTTCACTTTTGGAACTCTT	360												
QY	1848	TCCTGCACTGTGAGACTTCTTATAATCCCGAGAACTTTACAGCAAACTTCAGACTGTGCGCA	1907												
DB	361	TCCTGCACTGTGAGACTTCTTATAATCCCGAGAACTTTACAGCAAACTTCAGACTGTGCGCA	420												
QY	1908	GAATTGTTCTGGAAGCCCTTCTGATTCAGAGAAGGATTTGTCTGGCAAGTCTGATCTAG	1967												
DB	421	GAATTGTTCTGGAAGCCCTTCTGATTCAGAGAAGGATTTGTCTGGCAAGTCTGATCTAG	480												
QY	1968	AGAATTCCTCCAGTCTGGAAGCCCTTCTGAGACCCCTGAGCATAGTTCTGGGGAGGAAG	2027												
DB	481	AGAATTCCTCCAGTCTGGAAGCCCTTCTGAGACCCCTGAGCATAGTTCTGGGGAGGAAG	540												
QY	2028	ATGACTGGGAATCTAGTGCAAGTGAAGCAGAGAGTCTCAAACTGTGAACTCATTCGTGTA	2087												
DB	541	ATGACTGGGAATCTAGTGCAAGTGAAGCAGAGAGTCTCAAACTGTGAACTCATTCGTGTA	600												
QY	2088	ATTCTGATGACCCCTACAACCCCTTTAAATTTTAAAGGCTCCCTTTTCAAACATCAGGGGAAA	2147												
DB	601	ATTCTGATGACCCCTACAACCCCTTTAAATTTTAAAGGCTCCCTTTTCAAACATCAGGGGAAA	660												
QY	2148	ATGAGAAAGGCTGTCTGACTCAAGAACCCTCATCTGAGTCGATTTGTGGCCATTTCTGAGT	2207												
DB	661	ATGAGAAAGGCTGTCTGACTCAAGAACCCTCATCTGAGTCGATTTGTGGCCATTTCTGAGT	720												
QY	2208	GTCAACCTTACTTTCTTTGTAAGGTGAGCTGTTTGGGGAGCC-AAGAAAGTGAATGTCCA	2266												
DB	721	GTCAACCTTACTTTCTTTGTAAGGTGAGCTGTTTGGGGAGCC-AAGAAAGTGAATGTCCA	780												
QY	2267	GACTTCGGTACAGCGTGACGTTCTTTCTGGAGGAAGACACACATGTCAAAAAGAAAAAG	2326												
DB	781	GACTTCGGTACAGCGTGACGTTCTTTCTGGAGGAAGACACACATGTCAAAAAGAAAAAG	840												
QY	2327	GTAA-CCTTCT													

and NotI enzymes and cloned into Sali/NotI site of  
pCMV-SPORT6 plasmid vector. The average insert size is  
about 3.6kb."

ORIGIN		Query Match		Score 803;		DB 6;		Length 835;	
Best Local Similarity		99.4%;		Pred. No. 1.1e-214;					
Matches 806;		Conservative 0;		Mismatches 5;		Indels 0;		Gaps 0;	
QY	339	CGGATCCGCTCGACCTCCAGCGCTGAGCGCGGGCCCTACCTGAGAGACTGTCAAGA	398						
DB	25	CGGGATATCGTCGACCTCCAGCGCTGAGCGCGCGCCCTACCTGAGAGACTGTCAAGA	84						
QY	399	AAAGGAGATGAGCGCGGGGACAGCGGATCGCGAAACGCTTGCCCTCGGGCGGCT	458						
DB	85	AAAGGAGATGAGCGCGGGGACAGCGGATCGCGAAACGCTTGCCCTCGGGCGGCT	144						
QY	459	TCCGGTCTCGGCCACCCCTTTTCCCTCGCGGATCGCAAGCAGGCTCTTTAAAGTTCGGA	518						
DB	145	TCCGGTCTCGGCCACCCCTTTTCCCTCGCGGATCGCAAGCAGGCTCTTTAAAGTTCGGA	204						
QY	519	CGCCTCTTGGCCCGGAAACTCCGGGAACCCACACTGCTTCTCTGCCAGCCCGAGA	578						
DB	205	CGCCTCTTGGCCCGGAAACTCCGGGAACCCACACTGCTTCTCTGCCAGCCCGAGA	264						
QY	579	CTCGGCTCAGTTACTGACGAAACTGCTCTCCAGCTCTTTGCGCGCTCCCGGATTCG	638						
DB	265	CTCGGCTCAGTTACTGACGAAACTGCTCTCCAGCTCTTTGCGCGCTCCCGGATTCG	324						
QY	639	TTCAAGAGGTGCTAATTTTGGAGCCAACTTTTTCGGTGGAAATGTTTCGACACAGATGGCTAG	698						
DB	325	TTCAAGAGGTGCTAATTTTGGAGCCAACTTTTTCGGTGGAAATGTTTCGACACAGATGGCTAG	384						
QY	699	ATTTTGTGGAGTCTACAGCGCCCTGAGAGCCCTGAGGGACGGAGAACCCGCCGCC	758						
DB	385	ATTTTGTGGAGTCTACAGCGCCCTGAGAGCCCTGAGGGACGGAGAACCCGCCGCC	444						
QY	759	CCACAGCGCAGAAATCTTTGAGTTGCTGCGAGCTGACTCTCTCAGACCCCTCGGTCAACA	818						
DB	445	CCACAGCGCAGAAATCTTTGAGTTGCTGCGAGCTGACTCTCTCAGACCCCTCGGTCAACA	504						
QY	819	GTCCCTCTTGAATGAGGAGGGATCCATCGGCAATATCTCGCCGCCAGACCTAAAT	878						
DB	505	GTCCCTCTTGAATGAGGAGGGATCCATCGGCAATATCTCGCCGCCAGACCTAAAT	564						
QY	879	TGGAGCTTAAGCCCAAGGAGTCTTTGGACCCTGCGACGACGCTTTTCTCTTAGAGC	938						
DB	565	TGGAGCTTAAGCCCAAGGAGTCTTTGGACCCTGCGACGACGCTTTTCTCTTAGAGC	624						
QY	939	AGCAGCTGTGGGAGTGGAGCTGTTGGCCAGTAGCCTTCAATCCCGTCTGTACTCTAAC	998						
DB	625	AGCAGCTGTGGGAGTGGAGCTGTTGGCCAGTAGCCTTCAATCCCGTCTGTACTCTAAC	684						
QY	999	GGGAACCTTGGCTCTTTCGCCCTCTGGGCCCTCTAAACATTCACCGATAGACAAATTCAGTG	1058						
DB	685	GGGAACCTTGGCTCTTTCGCCCTCTGGGCCCTCTAAACATTCACCGATAGACAAATTCAGTG	744						
QY	1059	TGGTATCTCTATTTGCTGAAACCTTCTACCTGGAAGTCTTCTTAGGCTAGAGTCAAGT	1118						
DB	745	TGGTATCTCTATTTGCTGAAACCTTCTACCTGGAAGTCTTCTTAGGCTAGAGTCAAGT	804						
QY	1119	ATCAGAACAGTGTGGAATAGCGAGGTAGT	1149						
DB	805	ATCAGAACAGTGTGGAATAGCGAGGTAGT	835						

RESULT 8  
BM559080  
LOCUS  
DEFINITION AGNCOURT\_6563011 NIH\_MGC\_67 Homo sapiens cDNA clone IMAGE:5551220  
5', mRNA sequence.  
ACCESSION BM559080  
VERSION BM559080.1 GI:18802370

KEYWORDS		EST.	
SOURCE		Homo sapiens (human)	
ORGANISM		Homo sapiens	
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
AUTHORS		Mammalia; Rutheria; Primates; Catarrhini; Hominidae; Homo.	
TITLE		NIH-MGC http://mgi.nci.nih.gov/.	
JOURNAL		National Institutes of Health, Mammalian Gene Collection (MGC)	
COMMENT		Unpublished (1999)	
		Contact: Robert Strausberg, Ph.D.	
		Email: cgapbs-xemail.nih.gov	
		Tissue Procurement: ATCC	
		cDNA Library Preparation: Life Technologies, Inc.	
		cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)	
		DNA Sequencing by: Agencourt Bioscience Corporation	
		Clone distribution: MGC clone distribution information can be	
		found through the I.M.A.G.E. Consortium/LLNL at:	
		http://image.llnl.gov	
		Plate: L1AM12264 row: j column: 21	
		High quality sequence stop: 599.	
FEATURES		Location/Qualifiers	
source		1..1021	
		/organism="Homo sapiens"	
		/mol_type="mRNA"	
		/db_xref="taxon:9606"	
		/clone="IMAGE:5551220"	
		/tissue_type="retinoblastoma"	
		/lab_host="DH10B (phage-resistant)"	
		/clone_lib="NIH_MGC_67"	
		/notes="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;	
		Site_2: Sali; Cloned unidirectionally. Primer: Oligo dT.	
		Average insert size 1.75 kb. Library constructed by Life	
		Technologies."	
ORIGIN		Query Match	
		Best Local Similarity	
		26.4%; Score 777.2; DB 4; Length 1021;	
		93.3%; Pred. No. 2.3e-207;	
		Matches 889; Conservative 0; Mismatches 48; Indels 16; Gaps 7;	
QY	1219	TGCAAGACTCATTCGGCCCTCGTGGCAGGATGTCACACCTCTTCTACGGAAGCCCTACC	1278
DB	1	TGCAAGACTCATTCGGCCCTCGTGGCAGGATGTCACACCTCTTCTACGGAAGCCCTACC	60
QY	1279	AGAAATTCACCATCTTCGCATGAAACGGCTGGAATTCCTTCAACAGGCTAACAGGGGCA	1338
DB	61	AGAAATTCACCATCTTCGCATGAAACGGCTGGAATTCCTTCAACAGGCTAACAGGGGCA	120
QY	1339	AGATTATCCACCCCTGACAGGATATGGCTACACAGCTGGAGGAGGAACACAGCCT	1398
DB	121	AGATTATCCACCCCTGACAGGATATGGCTACACAGCTGGAGGAGGAACACAGCCT	180
QY	1399	TCTCCGATGGATCCAAACACATGCGAGATAACCCCAACAGTTTCTCTCTGCTGG	1458
DB	181	TCTCCGATGGATCCAAACACATGCGAGATAACCCCAACAGTTTCTCTCTGCTGG	240
QY	1459	AGACATTCCTGGAAACACCCAGGAATCCACTGAAGAAAAAATAGAATTTAACTACAGA	1518
DB	241	AGACATTCCTGGAAACACCCAGGAATCCACTGAAGAAAAAATAGAATTTAACTACAGA	300
QY	1519	GGTTCCATCTGCTTTGGAAGAGAGCCCTTCTGAGGCTGTCCATCTAGTAGATACC	1578
DB	301	GGTTCCATCTGCTTTGGAAGAGAGCCCTTCTGAGGCTGTCCATCTAGTAGATACC	360
QY	1579	TATGAAAAAGGAGCTGGAGAGGGCCGAATAAGTGTAGTTGATTACTCATACCTAGGAAG	1638
DB	361	TATGAAAAAGGAGCTGGAGAGGGCCGAATAAGTGTAGTTGATTACTCATACCTAGGAAG	420
QY	1639	TGACCTTTCCATTTCTGCCAGACAGCTTGTAGTAACAACTGATAGATTATATTTGGG	1698
DB	421	TGACCTTTCCATTTCTGCCAGACAGCTTGTAGTAACAACTGATAGATTATATTTGGG	480
QY	1699	AGGTGATTCAGTCACTTGGAAACAGTTCTGTATCCAGAGGTGAGGATGGGATGAGGA	1758



Db 481 AGGTGCATCCAGTGACCTGGAAACAAGTTCTGTATCCAGAGGTGAGGATGGGATGAGGA 540  
Qy 1759 AGCTGAGGATGATGGTTTTCATAGTAGTAGTCACTGTCTAGACTCAGACTCAGACCTTGAACAAGA 1818  
Db 541 AGCTGAGGATGATGGTTTTCATAGTAGTAGTCACTGTCTAGACTCAGACTCAGACCTTGNACAGA 600  
Qy 1819 CCCTGAAGGGCTTCACTTTGGAACTCTTTCTGAGGTGTAGATCCTTATATATCCCAAGAA 1878  
Db 601 CCCTGAAGGGCTTCACTTTGGAACTCTTTCTGAGGTGTAGATCCTTATATATCCCAAGAA 660  
Qy 1879 CTTTACAGCACAACTCAGACTGCTGCAGAAATGTTCTCTG-AAGAGCCTTCTGTATTCAG 1937  
Db 661 CTTTACAGCACAACTCAGACTGCTGCAGAAATGTTCTCTGAAAGAGCCTTCTGTATTCAG 720  
Qy 1938 AGAAGGATTTGTCTGGCAAGTCTGATCTAGAGAAATCTCTCCAGTCTGGAGCCTTCTCTG 1997  
Db 721 AGAAGGATTTGTCTGGCAGTCTGATCTAGAGAAATCTCTCCANTCTGGAGCCTTCCNN 780  
Qy 1998 AGA---CCCTGAGCAGTAGTT-CTGGGGAGGAAGA--TGACTGGGAATCTAGTGAGATG 2051  
Db 781 TGAGACCCCTGAGCAGTAGTTCTCTGGGGAGGAAGAATGACTGGNGAATCTAGTGAGATG 840  
Qy 2052 AAGCAG-AGAGTCTCAACTGTGGAATCTATCTGTAAATCTGTATGTA-----CCCTACA 2105  
Db 841 AAGCANAAGAGTCTCAACTGTGGAATCTATTTCTGTAAATCTGGATGACCCCTTAACA 900  
Qy 2106 ACCCTTTAAA---TTTTAAGGCTCTTTTCAAAATCAGGGGAAATGAGAAA 2155  
Db 901 ACCCTTTAAAATTTTAAAGGCTCTTTTTCNAATCTCAGGGGGAAGAAA 953

RESULT 9  
CD642687  
LOCUS  
DEFINITION AGENCOURT\_14569880 NIA Human H1 Embryonic Stem Cell cDNA Library  
(Long) Homo sapiens cDNA clone IMAGE:30427369 5', mRNA sequence.  
CD642687  
CD642687.1 GI:31812817  
EST.  
Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 775)  
NIA-MGC <http://mgc.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Daniela S. Gerhard, Ph.D.  
Office of Cancer Genomics  
National Cancer Institute / NIH  
Bldg. 31 Rm10A07 Bethesda, MD 20892  
Email: [cgapbe@mail.nih.gov](mailto:cgapbe@mail.nih.gov)  
Tissue Procurement: Irene Ginis and Mahendra Rao, NIA  
cDNA Library Preparation: Yulan Piao and Minoru KO  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC c lone distribution information  
can be found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: NDAM514 row: 1 column: 02  
High quality sequence stop: 730.  
Location/Qualifiers  
1. 775  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:30427369"  
/tissue type="Embryonic stem cells"  
/cell line="WA01"  
/lab\_hosts="DH10B (T1 phage-resistant)"  
/clone\_libs="NIA Human H1 Embryonic Stem Cell cDNA Library  
(Long)"  
/note="Vector: pCMV-Sport6; Site\_1: NotI; Site\_2: SalI;

## ORIGIN

Query Match 26.1%; Score 768.2; DB 6; Length 775;  
Best Local Similarity 99.4%; Pred. No. 7.1e-205;  
Matches 770; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
Qy 356 CCAGCGCTGAGCGCGCGCCCTACTGAGAGACTGTCAAGAAAAGGAGATGGAGCCG 415  
Db 1 CCAGCGCTGAGCGCGCGCCCTACTGAGAGACTGTCAAGAAAAGGAGATGGAGCCG 60  
Qy 416 GGACAGGGCGGATCGCGGAAACGGCTTGGCCCTCGGGCGGCTTCGGTTCTGCCACCC 475  
Db 61 GGACAGGGCGGATCGCGGAAACGGCTTGGCCCTCGGGCGGCTTCGGTTCTGCCACCC 120  
Qy 476 TTTTTCCTCGCGGATCGCAAGCAGGCTCTTCTTAAGTTCGCCGCGCTCTTGGCCCGGAA 535  
Db 121 TTTTTCCTCGCGGATCGCAAGCAGGCTCTTCTTAAGTTCGCCGCGCTCTTGGCCCGGAA 180  
Qy 536 AACTCGGGAAACCCCACTGCTTTCTGCGCCAGCCGAGACTCGGGTCACTTACTGG 595  
Db 181 AACTCGGGAAACCCCACTGCTTTCTGCGCCAGCCGAGACTCGGGTCACTTACTGG 240  
Qy 596 ACCAAGTCTCTCCAGCTCTTGGCGCGCTCCCGGATTCCTCAGAGGCTGCTAATT 655  
Db 241 ACCAAGTCTCTCTCCAGCTCTTGGCGCGCTCCCGGATTCCTCAGAGGCTGCTAATT 300  
Qy 656 TGGAGCAACATTTTCGGTGGAAATGTTTCGACACAGATGGCTAGATTTTGTGGAGTCTAC 715  
Db 301 TGGAGCAACATTTTCGGTGGAAATGTTTCGACACAGATGGCTAGATTTTGTGGAGTCTAC 360  
Qy 716 AGCGCCTGAGAGCCTTGAAAGGAGCGGAGAAAACAGCCGCCCCCAAGCGAGCAATCT 775  
Db 361 AGCGCCTGAGAGCCTTGAAAGGAGCGGAGAAAACAGCCGCCCCCAAGCGAGCAATCT 420  
Qy 776 TTGAGTTCGCTGAGCTCGACTCTCTCAGACCCCTCGGTCAACAGTCCCTTGTGCTA 835  
Db 421 TTGAGTTCGCTGAGCTCGACTCTCTCAGACCCCTCGGTCAACAGTCCCTTGTGCTA 480  
Qy 836 GAGGAGGGGATCCACTTGGCAATACTCGCCGCCCAAGACTTAAATTTGGAGCTTAAAGGCCAAG 895  
Db 481 GAGGAGGGGATCCACTTGGCAATACTCGCCGCCCAAGACTTAAATTTGGAGCTTAAAGGCCAAG 540  
Qy 896 GGAAGTGTCTTGGACCCCTGACAGAGGCTTTTCTTTAGAGCAGCAGCTGTGGGAGTG 955

This is a long-transcript enriched cDNA library (Genome Res. 11: 1553-1558 (2001)). [PMID: 11544199] from WA01 cell line. Undifferentiated human ES cell line WA01/H1 was obtained from WiCell Research Institute, Inc., Madison, WI, cultured according to their instructions, on MEF feeders. They formed round colonies with defined edges and were positive for alkaline phosphatase, SSEA-4, OCT3, OCT4, REX1, UTP, TERT, SOX2, CX43 and CX45. They are negative for GATA2, GATA4, PDX1, NCAM, MSX1, FLT3, SSEA-1, TUBB3, NES, GFAP, and EOMES. When confluent (18-10 days after plating), the ES cells from 4 X 6cm dishes were treated with 1 mg/ml collagenase, type IV (Invitrogen/GIBCO) for 5-10 min and gently scraped off with 5 ml pipette. RNA was purified with TRIzol Reagent from Invitrogen. Protocol ref: Genome Res. 11: 1553-1558 (2001). [PMID: 11544199] Double-stranded cDNAs were synthesized with an Oligo(dT) primer [invitrogen: 5'-pGACTAGTTCAGTCGAGCGCGCCCTTTTTCCTTTT-3'] from 3.4g of total RNA, treated with T4 DNA polymerase, and purified by ethanol-precipitation. The cDNAs were ligated to Lone-linker LL-Sal4, purified by phenol/chloroform extraction, and separated from free linkers by Centricon-100 column. Then, the cDNAs were amplified by long-range high fidelity PCR using Ex Taq polymerase (Takara) with a primer Sal4-S for 25 cycles. The products were purified by phenol/chloroform extraction and Centricon-100 column. The cDNAs were digested with SalI and NotI enzymes and cloned into SalI/NotI site of pCMV-Sport6 plasmid vector. The average insert size is about 3.6kb.



Db 541 GGAAGTCTTTGGACCTCGACGACAGGCTTTCTCTTAGACGACGCTGTGGGAGTG 600  
Qy 956 GAGCTGTTGCCAGTAGCCTCAATCCCGTCTGTACTCTTAACCGGAACTTGGCTCTTCG 1015  
Db 601 GAGCTGTTGCCAGTAGCCTCAATCCCGTCTGTACTCTTAACCGGAACTTGGCTCTTCG 660  
Qy 1016 CCCTCTGGGCTCTAAACATTCAACGCATAGACAATTTTCAGTGTGGTATCCTATTTCGTG 1075  
Db 661 CCCTCTGGGCTCTAAACATTCAACGCATAGACAATTTTCAGTGTGGTATCCTATTTCGTG 720  
Qy 1076 AACCTTCTCTACTGAGCTGCTTTCTTAGCTAGGCTAGAGTCAAGTATACGAAAGTG 1130  
Db 721 AACCTTCTCTACTGAGCTGCTTTCTTAGCTAGGCTAGAGTCAAGTATACGAAAGTG 775

RESULT 10  
CA749002/c  
LOCUS  
DEFINITION  
UI-H-FEI-bei-a-04-0-UI.s2 NCI CGAP FE1 Homo sapiens cDNA clone  
UI-H-FEI-bei-a-04-0-UI 3', mRNA sequence.  
ACCESSION  
CA749002  
VERSION  
CA749002.1 GI:25569930  
KEYWORDS  
EST.  
Homo sapiens (human)  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 793)  
NCI-CCGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-re@mail.nih.gov](mailto:cgapbs-re@mail.nih.gov)  
Tissue Procurement: James Martin  
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Clone distribution information can be obtained  
from Dr. M. Bento Soares, [bento-soares@uiowa.edu](mailto:bento-soares@uiowa.edu)  
The following repetitive elements were found in this cDNA  
sequence: 1-22, >AT rich#Low\_complexity  
Seq primer: M13 FORWARD  
POLYA=Yes.

FEATURES  
source  
1..793  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="UI-H-FEI-bei-a-04-0-UI"  
/tissue\_type="Cell lines"  
/dev\_stage="Adult"  
/lab\_host="DH10B (Life Technologies)"  
/clone\_lib="NCI CGAP FE1"  
/note="Organ: Chondrosarcoma; Vector: p773-Pac  
(Pharmacia) with a modified polylinker; Site 1: Ecor I;  
Site 2: Not I; NCI CGAP FE1 is a normalized cDNA library  
derived from a pool of mRNA obtained from 3 cell lines  
from Grade II chondrosarcoma tissues. The library was  
constructed according to Bonaldo, Lennon and Soares,  
Genome Research, 6:791-806, 1996. First strand cDNA  
synthesis was primed with an oligo-dT primer containing a  
Not I site. Double stranded cDNA was ligated to an Ecor I  
adaptor, digested with Not I, and cloned directionally  
into p773-Pac vector. The oligonucleotide used to prime  
the synthesis of first-strand cDNA contains a library tag  
sequence that is located between the Not I site and the  
(dT)18 tail. The sequence tag for this library is  
CGCTACGGAC. The cell lines were provided by Dr James  
Martin from the University of Iowa.  
TAG TISSUE=Human grade 2 chondrosarcoma cell line pool  
TAG\_Lib=UI-H-FEI

ORIGIN  
TAG\_SEQ=CGCTACGGAC"  
Query Match 25.8%; Score 758; DB 6; Length 793;  
Best Local Similarity 99.1%; Pred. No. 5.5e-202; Indels 0; Gaps 0;  
Matches 761; Conservative 0; Mismatches 7;  
Qy 1570 TGAGATACCTATGCGAAAGAGAGCCTGGAGAGGGCCGAATAAGTGTAGTGTACTACTCAT 1629  
Db 793 TGAGATACNTATGCGAAAGAGAGCCTGGAGAGGGCCGAATAAGTGTAGTGTACTACTCAT 734  
Qy 1630 CCTAGAAAGGTGACCTTCCCAATTTCTGCCAGACAGCTTGTAGTAAACAACTGATAGATTA 1689  
Db 733 CNTAGAAAGGTGACCTTCCCAATTTCTGCCAGACAGCTTGTAGTAAACAACTGATAGATTA 674  
Qy 1690 TATTTTGGGAGGTGCATCCAGTGACCTGGAAACAAAGTTCTGATCCACAAGGTGAGGATTTG 1749  
Db 673 TATTTTGGGAGGTGCATCCAGTGACCTGGAAACAAAGTTCTGATCCACAAGGTGAGGATTTG 614  
Qy 1750 GGATGAGAAAGCTGAGGATGATGGTTTTGTAGTAGTGTAGTGTACTGTGAGACTCAGACCT 1809  
Db 613 GGATGAGAAAGCTGAGGATGATGGTTTTGTAGTAGTGTAGTGTACTGTGAGACTCAGACCT 554  
Qy 1810 TGAACAAAGACCTGAAAGGGCTTCACTTTGGAACTCTTTCTGCAAGTGTAGATCCTTATAA 1869  
Db 553 TGAACAAAGACCTGAAAGGGCTTCACTTTGGAACTCTTTCTGCAAGTGTAGATCCTTATAA 494  
Qy 1870 TCCCCAGAACTTTTACAGCAACAATTCCAGCTGCTGCCAGAAATTTCTCTGAAGAGCCTTC 1929  
Db 493 TCCCCAGAACTTTTACAGCAACAATTCCAGCTGCTGCCAGAAATTTCTCTGAAGAGCCTTC 434  
Qy 1930 TGATTTCAGAGAAGGATTTGTCTGGCAAGTCTGATCTAGAGAAATTTCTCCAGTCTGGAAG 1989  
Db 433 TGATTTCAGAGAAGGATTTGTCTGGCAAGTCTGATCTAGAGAAATTTCTCCAGTCTGGAAG 374  
Qy 1990 CCTTCTGAGACCCCTGAGCATAGTTCTGGGGAGGAAGATGACTGGGAATCTAGTCAGA 2049  
Db 373 CCTTCTGAGACCCCTGAGCATAGTTCTGGGGAGGAAGATGACTGGGAATCTAGTCAGA 314  
Qy 2050 TGAAGCAGAGAGTCTCAAACTGTGGAACTCAATCTGTAAATCTGATGACCCCTACAAACC 2109  
Db 313 TGAAGCAGAGAGTCTCAAACTGTGGAACTCAATCTGTAAATCTGATGACCCCTACAAACC 254  
Qy 2110 TTTAAATTTTAAAGCTCCTTTTCAACATCAGGGGAAAAATGAGAAAGGCTGTGTGATC 2169  
Db 253 TTTAAATTTTAAAGCTCCTTTTCAACATCAGGGGAAAAATGAGAAAGGCTGTGTGATC 194  
Qy 2170 AAAGACCCCATCTGAGTCCATTTGTGGCCATTTCTGAGTGTGCACACCTTACTTTCTGTAA 2229  
Db 193 AAAGACCCCATCTGAGTCCATTTGTGGCCATTTCTGAGTGTGCACACCTTACTTTCTGTAA 134  
Qy 2230 GGTGCAGCTGTGGGGAGCCAAAGTGAATGTCCAGACTCGGTACAGCGTGCAGCTTCT 2289  
Db 133 GGTGCAGCTGTGGGGAGCCAAAGTGAATGTCCAGACTCGGTACAGCGTGCAGCTTCT 74  
Qy 2290 TTCTGGAGGAAGACACACATGTCAAAAGAAAAAGGTAACTTCT 2337  
Db 73 TTCTGGAGGAAGACACACATGTCAAAAGAAAAAGGTGTGTTCTT 26  
RESULT 11  
CA426770/c  
LOCUS  
DEFINITION  
UI-H-FEI-bei-a-04-0-UI.s1 NCI CGAP FE1 Homo sapiens cDNA clone  
UI-H-FEI-bei-a-04-0-UI 3', mRNA sequence.  
ACCESSION  
CA426770  
VERSION  
CA426770.1 GI:24789496  
KEYWORDS  
EST.  
Homo sapiens (human)  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 795)

AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: James Martin  
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Clone distribution information can be obtained  
from Dr. M. Bento Soares, bento-soares@uiowa.edu  
The following repetitive elements were found in this cDNA  
sequence: 1-22, >AT rich#Low\_complexity  
Seq primer: M13 FORWARD  
POLYA=Yes.

FEATURES  
source Location/Qualifiers  
1. .795  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="UI-H-FEI-bel-a-04-0-UI"  
/tissue\_type="Cell lines"  
/dev\_stage="Adult"  
/lab\_host="DH10B (Life Technologies)"  
/clone\_lib="NCI CGAP FE1"  
/notes="Organ: Chondrosarcoma; Vector: p7T3-Pac  
(Pharmacia) with a modified polylinker; Site 1: Ecor I;  
Site 2: Not I; NCI CGAP FE1 is a normalized cDNA library  
derived from a pool of mRNA obtained from 3 cell lines  
from grade II chondrosarcoma tissues. The library was  
constructed according to Bonaldo, Lennon and Soares,  
Genome Research, 6:791-806, 1996. First strand cDNA  
synthesis was primed with an oligo-dT primer containing a  
Not I site. Double stranded cDNA was ligated to an Ecor I  
adaptor, digested with Not I, and cloned directionally  
into p7T3-Pac vector. The oligonucleotide used to prime  
the synthesis of first-strand cDNA contains a library tag  
sequence that is located between the Not I site and the  
(dT)18 tail. The sequence tag for this library is  
CGTACGGAC. The cell lines were provided by Dr James  
Martin from the University of Iowa.  
TAG TISSUE=Human grade 2 chondrosarcoma cell line pool  
TAG LIB=UI-H-FEI  
TAG\_SEQ=CGTACGGAC"

ORIGIN  
Query Match 25.7%; Score 755.6; DB 6; Length 795;  
Best Local Similarity 98.8%; Pred. No. 2.6e-201;  
Matches 761; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1568 AGTGAGATACCTATGGAAGAGCCCTGGAGGCGCGAATAAGTGTAGTTGATTACTCA 1627  
DB |||||||  
795 AGTGAGATACCTATGGAAGAGAGCCCTGGAGGCGCGAATAAGTGTAGTTGATTACTCA 736  
QY 1628 TACCTAGAAGGTGACCTTCCCATTTCTGCCAGACCACTGTTAGTAAACAACTGATAGAT 1687  
DB |||||||  
735 TACCTAGAAGGTGACCTTCCCATTTCTGCCAGACCACTGTTAGTAAACAACTGATAGAT 676  
QY 1688 TATATTTTGGAGGTGATCCAGTGACCTGGAAACAAGTCTTGATCCAGAGGTGAGGAT 1747  
DB |||||||  
675 TATATTTTGGAGGTGATCCAGTGACCTGGAAACAAGTCTTGATCCAGAGGTGAGGAT 616  
QY 1748 TGGGATCAGGAAGCTGAGGATGATGGTTTGTATGATGATAGCTCAGTCCAGTCCAGCTCAGAC 1807  
DB |||||||  
615 TGGGATCAGGAAGCTGAGGATGATGGTTTGTATGATGATAGCTCAGTCCAGTCCAGCTCAGAC 556  
QY 1808 CTTGAACAAGACCTGAAGGCTTTCACCTTTTGGAACTCTTTCTGCAGGTAGATCCTTAT 1867  
DB |||||||  
555 CTTGAACAAGACCTGAAGGCTTTCACCTTTTGGAACTCTTTCTGCAGGTAGATCCTTAT 496  
QY 1868 AATCCCAGAACTTTACAGCAACAAATTCAGACTCTGCCAGAAATGTTCTCTGAAGACCT 1927  
DB |||||||

Db 495 AATCCCCAGAACTTTACAGCAACAATTCAGACTGCTGCCAGAAATGTTCTCTGAAGACCT 436  
QY 1928 TCTGATTCAGAGAAAGGATTTGTCTGCAAGTCTGTAGATCTAGAGAAATTCCTCCAGTCTGGA 1987  
DB |||||||  
435 TCTGATTCAGAGAAAGGATTTGTCTGCAAGTCTGTAGATCTAGAGAAATTCCTCCAGTCTGGA 376  
QY 1988 AGCTTCTCTGAGACCCCTGAGCATAGTTCTGGGAGGAAGATGACTGGGAATCTAGTGCA 2047  
DB |||||||  
375 AGCTTCTCTGAGACCCCTGAGCATAGTTCTGGGAGGAAGATGACTGGGAATCTAGTGCA 316  
QY 2048 GATGAAGCAGAGAGTCTCAAACTGTGGAACCTCAATCTGTAAATCTCATGACCCCTACAAC 2107  
DB |||||||  
315 GATGAAGCAGAGAGTCTCAAACTGTGGAACCTCAATCTGTAAATCTCATGACCCCTACAAC 256  
QY 2108 CCTTTAAATTTTAAAGGCTCTCTTTTCAACATCAGGGGAAAAATGAGAAAAGCTCTCTGAC 2167  
DB |||||||  
255 CCTTTAAATTTTAAAGGCTCTCTTTTCAACATCAGGGGAAAAATGAGAAAAGCTCTCTGAC 196  
QY 2168 TCAAGACCCCATCTGAGTCCATTTGTGGCCATTTCTGAGTGTACACCTTACTTTCTTGT 2227  
DB |||||||  
195 TCAAGACCCCATCTGAGTCCATTTGTGGCCATTTGTGAGTGTACACCTTACTTTCTTGT 136  
QY 2228 AGGTGCAGCTGTTGGGAGCCCAAGAAAGTGAATGTCCAGACTCGGTACAGCGTGACGTT 2287  
DB |||||||  
135 AGGTGCAGCTGTTGGGAGCCCAAGAAAGTGAATGTCCAGACTCGGTACAGCGTGACGTT 76  
QY 2288 CTTTCTGGAGGAGACACACATGTCAAAAGAAAAAGGTAACTTCTCT 2337  
DB |||||||  
75 CTTTCTGGAGGAGACACACATGTCAAAAGAAAAAGGTAACTTCTCT 26

RESULT 12  
BG721970/c 850 bp mRNA linear EST 08-MAY-2001  
LOCUS 602698720F1 NIH\_MGC\_97 Homo sapiens cDNA clone IMAGE:4830749 5',  
DEFINITION mRNA sequence.  
ACCESSION BG721970  
VERSION BG721970.1 GI:14001157  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 850)  
TITLE NIH-MGC <http://mgc.nci.nih.gov/>.  
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.  
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki  
Toshiyuki and Piero Carninci (RIKEN)  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LLAM10752 row: g column: 06  
High quality sequence stop: 809.  
Location/Qualifiers  
1. .850  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4830749"  
/lab\_host="DH10B"  
/clone\_lib="NIH\_MGC\_97"  
/notes="Organ: testis; Vector: pBluescriptR (modified  
pBluescript KS+); Site 1: BamHI; Site 2: SalI-XhoI  
(gtcgag); Oligo-dT primed using primer  
5'-TTTTTTTTTTTTTNN-3', size-selected for average  
insert size 2.2 kb and normalized to ROT 5. This is a  
primary library enriched for full-length clones and



```

481 CATTGTGCCATTTCTGAGTCTCACACTTACTTTCTTTGTAAGGTGACGCTGTGTGGGAG 540
2248 CCAGAAGTGAATGTCAGACTCGGTACAGCGTGTCTTCTTCTGAGGAGACACAC 2307
541 CCAAGAAAGTGAATGTCAGACTCGGTACAGCGTGTCTTCTTCTGAGGAGACACAC 600
2308 ACATGTCAAAAGAAAAAGGTAACTTCTTCTTGAAGAGTGTACTGAGTATTATATAAGTGG 2367
601 ACATGTCAAAAGAAAAAGGTAACTTCTTCTTGAAGAGTGTACTGAGTATTATATAAGTGG 660
2368 TGATGAGATCGAAAGACCATGGGAGAAATTTGCAAGGATGATGCGAGTTCCAGAA 2427
661 TGATGAGATCGCAAGAGCACCATGGGAGAAATTTGCAAGGATGATGCGAGTTTCAGA 720
2428 ACGAATTCAGAAACAGAGATGCTATTGG--AATGTTGTTGACATTTGAACACAGAGAA 2485
721 AACATTTTCAGAAACAGAGATGCTATTGGATTTGTTGATTTGCTTTTGGACCCGAGA 780
2486 AG--AATGTTTAAATAGACTCCAGGAA 2510
781 AGGAATGTTTATAAGACTCCAGGGGA 807

```

```

RESULT 14
CF594071
LOCUS
DEFINITION AGENCOURT_15624090 NIH_MGC_147 Homo sapiens cDNA clone
IMAGE:30528106 5', mRNA sequence.
ACCESSION CF594071
VERSION CF594071.1 GI:36348197
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Daniela S. Gerhard, Ph.D.
Office of Cancer Genomics
National Cancer Institute / NIH
Bldg. 31 Rm10A07 Bethesda, MD 20892
Email: cgapbs-re@mail.nih.gov
Tissue Procurement: Dr. Stefan Hansson
cDNA Library Preparation: Michael J. Brownstein (NHGRI) with help
and advice from Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: NDAM613 row: i column: 11
High quality sequence stop: 595.
Location/Qualifiers
1. 838
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:30528106"
/tissue_type="Human Placenta"
/lab_host="DH10B Tona"
/clone_lib="NIH_MGC_147"
/notes="Organ: placenta; Vector: pBluescriptR; Site_1:
all-XhoI; Site_2: BamHI; Oligo-dT primed using primer
5'-TTTTTTTTTTTTTTVN-3', size-selected for average
insert size 2.3 kb and normalized to ROT 5. This is a
primary library enriched for full-length clones and
constructed using the Cap-trapper method (Carninci, in
preparation). Library constructed by M. Brownstein
(NIH/NHGRI, National Institutes of Health). Note: This is
a NIH_MGC library."

```

```

FEATURES
source
1. 838
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:30528106"
/tissue_type="Human Placenta"
/lab_host="DH10B Tona"
/clone_lib="NIH_MGC_147"
/notes="Organ: placenta; Vector: pBluescriptR; Site_1:
all-XhoI; Site_2: BamHI; Oligo-dT primed using primer
5'-TTTTTTTTTTTTTTVN-3', size-selected for average
insert size 2.3 kb and normalized to ROT 5. This is a
primary library enriched for full-length clones and
constructed using the Cap-trapper method (Carninci, in
preparation). Library constructed by M. Brownstein
(NIH/NHGRI, National Institutes of Health). Note: This is
a NIH_MGC library."

```

## ORIGIN

```

Query Match 25.0%; Score 734.6; DB 7; Length 838;
Best Local Similarity 97.8%; Pred. No. 2.3e-195;
Matches 787; Conservative 0; Mismatches 14; Indels 4; Gaps 4;
Qy 67 GGCTTTTCAACCTCTGTCTAGTTCGGAAAAACCATCGCGAGGCGGTGGGGGACTCTCTATC 126
Db 28 GGCTTTTCAACCTCTGTCTAGTTCGGAAAAACCATCGCGAGGCGGTGGGGGACTCTCTATC 87
Qy 127 CATGTGTGTAAGCGTCGAGCGGACCTAGGGAACCTCTTCTCCCGCCAGAGATGGAAGTCGC 186
Db 88 CATGTGTGTAAGCGTCGAGCGGACCTAGGGAACCTCTTCTCCCGCCAGAGATGGAAGTCGC 147
Qy 187 ATCAGTCGCGCCCTATTGGCGCGGCTGTCTTCTCCCTGTGTCTTCTGCGCCGCTGCCGCAT 246
Db 148 ATCAGTCGCGCCCTATTGGCGCGGCTGTCTTCTCCCTGTGTCTTCTGCGCCGCTGCCGCAT 207
Qy 247 TCCTGCTCTCTGTGGCTTTCTGTGGCTTCTGGAAGATCGGCTTGGAGCAGCGACGCCACC 306
Db 208 TCCTGCTCTCTGTGGCTTTCTGTGGCTTCTGGAAGATCGGCTTGGAGCAGCGACGCCACC 267
Qy 307 GCTGGGCAAGGCGGAGACTCTGTAGGCTTCTCGGAATCCCGTCCGACCTCCAGCGCTGA 366
Db 268 GCTGGGCAAGGCGGAGACTCTGTAGGCTTCTCGGAATCCCGTCCGACCTCCAGCGCTGA 327
Qy 367 GCCTCGCGGCTTACCTGAGAGACTGTCAAGAAAAAGGAGATGGAGCCGGGACAGCGCG 426
Db 328 GCCTCGCGGCTTACCTGAGAGACTGTCAAGAAAAAGGAGATGGAGCCGGGACAGCGCG 387
Qy 427 ATCGCGAAACGCTTGTGGCTTGTGGCGGCTTCCGGTTCTGCGCCACCTTTTTCCTCG 486
Db 388 ATCGCGAAACGCTTGTGGCTTGTGGCGGCTTCCGGTTCTGCGCCACCTTTTTCCTCG 447
Qy 487 GCATCGCAGCAGGCTTCTTAAGTTCGAGCGCTCTTGGCCCGGAAAACTCCGGGAA 546
Db 448 GCATCGCAGCAGGCTTCTTAAGTTCGAGCGCTCTTGGCCCGGAAAACTCCGGGAA 507
Qy 547 CCCACACTGCTTTCTCTGCGGCGGCTTCTTAAGTTCGAGCGCTTCTGGGCGGAACTGCT 606
Db 508 CCCACACTGCTTTCTCTGCGGCGGCTTCTTAAGTTCGAGCGCTTCTGGGCGGAACTGCT 567
Qy 607 CTCCAGCTCTTGTGCGCGCTTCCCGGATGCTTTCAGAGAGTCTTAATTTGGAGCCTAACT 666
Db 568 CTCCAGCTCTTGTGCGCGCTTCCCGGATGCTTTCAGAGAGTCTTAATTTGGAGCCTAACT 627
Qy 667 TTTTGGTGGATGCTTTTCGACAGATGGCTAGATTTTGTGGAGTCTACAGCGCCCTGAG 726
Db 628 TTTTGGTGGATGCTTTTCGACAGATGGCTAGATTTTGTGGAGTCTACAGCGCCCTGAG 687
Qy 727 AGCCCTGAAGGGACGGAGAAACAGCGCGCCCGCCAGCGCAGAAAATCTTTTGAG-TTCGC 785
Db 688 AGCCCTGAAGGGACGGAGAAACAGCGCGCCCGCCAGCGCAGAAAATCTTTTGAGTTTGC 747
Qy 786 TGAGCTCGACTCTCTCAGACCCCTC-GGTCCACAGTCCCTTG-ATTGGCTAGAGAGG 843
Db 748 TGAGCTCGACTCTCTCAGACCCCTC-GGTCCACAGTCCCTTG-ATTGGCTAGAGAGG 807
Qy 844 G-ATCCACTGGCAATACTCGCCCCC 867
Db 808 GAATCCACTGGCAATACTCGCCCCC 832

```

## RESULT 15

```

CB989883
LOCUS
DEFINITION AGENCOURT_13903499 NIH_MGC_147 Homo sapiens cDNA clone
IMAGE:30341796 5', mRNA sequence.
ACCESSION CB989883
VERSION CB989883.1 GI:30284403
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 793)  
NIH-MGC <http://mgc.ncl.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
Tissue procurement: Dr. Stefan Hansson  
cDNA Library Preparation: Michael J. Brownstein (NHGRI) with help  
and advice from Piero Carninci (RIKEN)  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: NDAM373 row: n column: 13  
High quality sequence stop: 654.

## FEATURES

## Source

1..793  
Location/Qualifiers  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:30341796"  
/issue\_type="Human Placenta"  
/lab\_host="DH10B Tona"  
/clone\_lib="NIH MGC 147"  
/note="Organ: placenta; Vector: pBluescriptR; Site\_1:  
all-XhoI; Site\_2: BamH; Oligo-dr primed using primer  
5'-TTTTTTTTTTTTTNN-3', size-selected for average  
insert size 2.3 kb and normalized to ROT 5. This is a  
primary library enriched for full-length clones and  
constructed using the Cap-trapper method (Carninci, in  
preparation). Library constructed by M. Brownstein  
(NIH/NHGRI, National Institutes of Health). Note: This is  
a NIH\_MGC library."

## ORIGIN

Query Match 24.8%; Score 728.4; DB 6; Length 793;  
Best Local Similarity 97.9%; Pred. No. 1.3e-193;  
Matches 738; Conservative 0; Mismatches 16; Indels 0; Gaps 0;  
QY 1 ATTTTGGGCTTGGCTTCCACCGCACCGCGGCTTACCCAGTCTCTTCGGGTATCGCGTTG 60  
Db 28 ATTTTGGGCTTGGCTTCCACCGCACCGCGGCTTACCCAGTCTCTTCGGGTATCGCGTTG 87  
QY 61 CTCAGGGGCTTTTCAACCTCTGTGAGTGGGAAACCATCGCGAGGCGGTGGGGGACT 120  
Db 88 CTCAGGGGCTTTTCAACCTCTGTGAGTGGGAAACCATCGCGAGGCGGTGGGGGACT 147  
QY 121 CCTATCCATGTTGTTGAAGCGTCGAGCGGACTAGGGAACCTCTTCCCGCCAGGATGGA 180  
Db 148 CCTATCCATGTTGTTGAAGCGTCGAGCGGACTAGGGAACCTCTTCCCGCCAGGATGGA 207  
QY 181 AGTCGCATCAGTCGCGGCTATATGCGGGGCTGTTCTTCCCTGTGTTCTGCGCGCGCTG 240  
Db 208 AGTCGCATCAGTCGCGGCTATATGCGGGGCTGTTCTTCCCTGTGTTCTGCGCGCGCTG 267  
QY 241 CGGATTCGCTGCCCTCTGTGGCTTTTCTGTGGTTCGAAGATCGGCGCTGGAGCAGGAC 300  
Db 268 CCGCATTCGCTGCCCTCTGTGGCTTTTCTGTGGTTCGAAGATCGGCGCTGGAGCAGGAC 327  
QY 301 GCCACGCTGGGCAAGGCGGAGCTCTGTAGGCTTCTCCGAATCCCGTCCAGCTCCAGC 360  
Db 328 GCCACGCTGGGCAAGGCGGAGCTCTGTAGGCTTCTCCGAATCCCGTCCAGCTCCAGC 387  
QY 361 CGCTAGCGCGCGGCGCTTACTCTGAGACATGTCTAAGAAAAAGGAGATGGAGCGGGGAC 420  
Db 388 CGCTAGCGCGCGGCGCTTACTCTGAGACATGTCTAAGAAAAAGGAGATGGAGCGGGGAC 447  
QY 421 AGCGGGATCGGGGAAACGGCTTGGCCCTTCGGGCGGGCTTCGGCTTCGGCCACCTTTT 480  
Db 448 AGCGGGATCGGGGAAACGGCTTGGCCCTTCGGGCGGGCTTCGGCTTCGGCCACCTTTT 507

QY 481 CCCTCGCGGATCGCAAGCAGGCTCTTCTAAGTTCCCGACGCTCTTGGCCCGGAAAACCTC 540  
Db 508 CCCTCGCGGATCGCAAGCAGGCTCTTCTAAGTTCCCGACGCTCTTGGCCCGGAAAACCTC 567  
QY 541 CGGGAACCCACACTGCTTTCTCTGCGCCAGCCGAGACTCGGGTCAGTTTACTGACGAA 600  
Db 568 CGGGAACCCACACTGCTTTCTCTGCGCCAGCCGAGACTCGGGTCAGTTTACTGACGAA 627  
QY 601 ACTGCTCTCCAGTCTCTTGGCGGCTCCCGGATTCCTTCAGNAGGTGCTTAATTTGGAG 660  
Db 628 ACTGCTCTCCAGTCTCTTGGCGGCTCCCGGATTCCTTCAGNAGGTGCTTAATTTGGAG 687  
QY 661 CCAACTTTTTCGGTGGAAATGTTTCGACACAGATGGCTAGATTTTGTCTGGAGTCTACAGGC 720  
Db 688 CCAACTTTTTCGGTGGAAATGTTTCGACACAGATGGCTAGATTTTGTCTGGAGTCTACAGGC 747  
QY 721 CCTGAGAGCCCTGAAGGACGGGAGAAACGAGC 754  
Db 748 CCTGAGAGCCCTGAAGGACGGGAGAAACGAGC 781

Search completed: September 16, 2005, 08:56:04  
Job time : 8799.62 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 15, 2005, 16:13:49 ; Search time 70.2388 Seconds  
(without alignments)  
3926.035 Million cell updates/sec

Title: US-10-650-482-2

Perfect score: 3804

Sequence: 1 MEEGTGSRKRLGPRAGFR.....RMFNRLQGTGCKGLNLVKQC 713

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : A\_Geneseq\_16Dec04:\*

1: Geneseqp1980s:\*

2: Geneseqp1990s:\*

3: Geneseqp2000s:\*

4: Geneseqp2001s:\*

5: Geneseqp2002s:\*

6: Geneseqp2003as:\*

7: Geneseqp2003bs:\*

8: Geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3804	100.0	713	8	AQ88458 Human GAD
2	3794	99.7	713	4	AAB95876 Human pro
3	3627	95.3	707	3	AAB53401 Human col
4	2223	58.4	698	8	AQ88460 Mouse GAD
5	813	21.4	153	4	AM25448 Human pro
6	260.5	6.8	674	3	AA184366 Human pro
7	260.5	6.8	674	7	ADE54748 Human pro
8	260.5	6.8	674	7	ADD45119 Human pro
9	260.5	6.8	674	7	ADE54744 Human pro
10	259.5	6.8	674	4	AAB92888 Human pro
11	259.5	6.8	674	8	ADR14089 Human pro
12	252.5	6.6	674	2	AAW98991 Human pro
13	238	6.3	657	2	AA141102 Human pro
14	237	6.2	657	5	AB57373 Mouse lsc
15	219	5.8	578	2	AAW79958 Human pro
16	219	5.8	578	2	AAW06514 Human pro
17	219	5.8	578	2	AA141104 Human pro
18	201	5.3	590	2	AA141101 Human pro
19	170	4.5	995	8	ADR98899 Lung spec
20	168	4.4	1454	8	ABM84777 Human dia
21	167	4.4	1596	6	ABG73372 Rat full
22	167	4.4	1596	7	ADE62326 Rat Prote
23	165	4.3	1596	2	AAW31347 Human pro
24	163	4.3	1500	8	ADM42998 Human INS
25	160.5	4.2	1562	4	ABG27493 Novel hum

Aae32723	KIAA0322
Abp97171	Human NED
Abp98330	Amino aci
Aam33192	Human pol
Aam40978	Human pol
Abg08505	Novel hum
Abb67865	Drosophil
Abb68397	Drosophil
Adi16260	Human nuc
Adf08473	Rat PAPIN
Adi38113	Human apo
Aaw48845	Human rec
Adi38115	Human H2O
Adp55733	Human PRO
Abg23910	Novel hum
Abg13903	Novel hum
Adp30190	Human sec
Adp30259	Human sec
Abi14747	Novel hum
Abp98840	Human str
Abi07120	Novel hum
Adi28089	Human nuc
Aao17706	Human PPA
Abg08504	Novel hum
Adm05088	Human PRO
Abi40294	Human ORF
Adg10872	Human STA
Aau30116	Novel hum
Adi71149	Human int
Aay53970	Human per
Adj70367	Human hea
Adi19538	Human PRO
Abg93245	C. albica
Adm05820	Human pro
Abg91810	Human int
Abb59642	Drosophil
Abw81546	Tumour re
Abu08929	Human tum
Adi25518	Binding d
Adm18025	MAGE-cl p
Adm33107	Human tum
Adi79399	Human MAG
Abg24167	Novel hum
Aay58634	Protein r
Adm42940	Human INS
Adj68723	Human hea
Abg95238	Human pro
Adk71860	Human kin
Adq97927	Mouse can
Abg95229	Human pro
Aam94010	Human sto
Adr09871	Human pro
Adi10462	Human pro
Ado85780	Rac/axin/
Abb05681	Human dif
Adi14035	Human NF-
Abm80164	Tumour-as
Abb62045	Drosophil
Abm83955	Human dia
Abb97562	Novel hum
Abp98339	Amino aci
Abg93152	S. cerevi
Aau30115	Novel hum
Aam79120	Human pro
Adm06070	Human pro
Adi18754	Human dia
Aam40114	Human pol
Abp96961	Human BMC
Adj49897	Rat TAO2
Adj58785	Rat TAO2
Abb61038	Drosophil
Abg02138	Novel hum
Aay66187	Human bla

99 140 3.7 685 8 ADR09563 Human pro  
100 139.5 3.7 635 4 ABG22101 Novel hum

ALIGNMENTS

RESULT 1  
ADQ88458  
ID ADQ88458 standard; protein; 713 AA.  
XX  
AC ADQ88458;  
XX  
DT 07-OCT-2004 (first entry)  
XX  
DE Human GADD34-like (GADD34L) protein.  
XX  
KW Oxidative stress; neuronal ischaemia; heart ischaemia; renal damage;  
KW autoimmune disease; neurodegenerative disorder; therapy; GADD34-like;  
KW GADD34L; eIF2alpha-specific regulatory subunit; phosphatase; human.  
XX  
OS Homo sapiens.  
XX  
PN US2004142345-A1.  
XX  
PD 22-JUL-2004.  
XX  
PF 28-AUG-2003; 2003US-00650482.  
XX  
PR 06-SEP-2002; 2002US-0408679P.  
XX  
PA (ROND/) RON D.  
PA (JOUS/) JOUSSE C.  
XX  
PI Ron D, Jousse C;  
XX  
DR WPI; 2004-552556/53.  
DR N-PSDB; ADQ88457.  
XX  
PT Screening test substances for preventing or treating disease involving  
PT oxidative stress, by testing test substances for its ability to inhibit  
PT activity of GADD34L and identifying test substance that inhibits activity  
PT of GADD34L.  
XX  
PS Disclosure; SEQ ID NO 2; 30pp; English.  
XX  
CC The present invention relates to a method of screening several test  
CC substances for preventing or treating diseases involving oxidative stress  
CC such as neuronal ischaemia, heart ischaemia, renal damage induced by  
CC ischaemia or toxins, autoimmune diseases and neurodegenerative disorders.  
CC The method involves testing the test substances for its ability to  
CC inhibit the activity of GADD34-like (GADD34L), also referred to as  
CC eIF2alpha-specific regulatory subunit of phosphatase, and identifying the  
CC test substance which inhibits the activity of GADD34L. The present  
CC sequence is the human GADD34L protein.  
XX  
SQ Sequence 713 AA;

Query Match 100.0%; Score 3804; DB 8; Length 713;  
Best Local Similarity 100.0%; Pred. No. 7.5e-303;  
Matches 713; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MPEGTGSRKRLGPRAGFRFPFPPRRSQAGSKFPTPLGPNSTLLSSAQPTRV 60  
DB 1 MPEGTGSRKRLGPRAGFRFPFPPRRSQAGSKFPTPLGPNSTLLSSAQPTRV 60  
QY 61 SYWTKLLSQLAPLPGLLQKVLINSQLFGGMFPTRWLDVAGVYALRALKGREKPAAPTA 120  
DB 61 SYWTKLLSQLAPLPGLLQKVLINSQLFGGMFPTRWLDVAGVYALRALKGREKPAAPTA 120  
QY 121 QKSLSSQLDSSDPSTVPLWLEEGIHQVSPDLEKLEKAGSALDPAQAFLLEQOL 180  
DB 121 QKSLSSQLDSSDPSTVPLWLEEGIHQVSPDLEKLEKAGSALDPAQAFLLEQOL 180

QY 181 WGVLLPSSLSQRLYSNRELSSPSGLNQRIDNFSVSVYLLNPSYLDPCPRLEVSQN 240  
DB 181 WGVLLPSSLSQRLYSNRELSSPSGLNQRIDNFSVSVYLLNPSYLDPCPRLEVSQN 240  
QY 241 SDGNSEVVGFTLTPESSCLREDHCHPOPLSAELIPASWQCPCPLSTEGLEPEIHLRMKR 300  
DB 241 SDGNSEVVGFTLTPESSCLREDHCHPOPLSAELIPASWQCPCPLSTEGLEPEIHLRMKR 300  
QY 301 LEFLQOANKGQDLPTPDQNGYHSLLEHSLLRMDPKHCRDNPTQFVPAAGDIIFGNTOES 360  
DB 301 LEFLQOANKGQDLPTPDQNGYHSLLEHSLLRMDPKHCRDNPTQFVPAAGDIIFGNTOES 360  
QY 361 TEEKIELLTTEVPPLALEEESPSECPSPSEIPMEKEPGEGRISVVVDYSYLEGDLFISARPA 420  
DB 361 TEEKIELLTTEVPPLALEEESPSECPSPSEIPMEKEPGEGRISVVVDYSYLEGDLFISARPA 420  
QY 421 CSNKLDIYILGGASSDLETSDDPEGEDWDEAEADGDFDSSLSLSDSDLEQDPEGLHLWNS 480  
DB 421 CSNKLDIYILGGASSDLETSDDPEGEDWDEAEADGDFDSSLSLSDSDLEQDPEGLHLWNS 480  
QY 481 FCSVDYPNPNFTATIQTAARIIVPEEPSDESEKOLSGKSDLENSQSGLPETPHSSGEE 540  
DB 481 FCSVDYPNPNFTATIQTAARIIVPEEPSDESEKOLSGKSDLENSQSGLPETPHSSGEE 540  
QY 541 DWESSADEAESLKLWNSFCNSDDPYNPLNFKAPFOTSGENEKGRDSTKTPSESIVASE 600  
DB 541 DWESSADEAESLKLWNSFCNSDDPYNPLNFKAPFOTSGENEKGRDSTKTPSESIVASE 600  
QY 601 CHTLLCKVOLLSQSESECPDSVQORDVLSGGRHTRVKRKKVTFLEEVTEYIISGDEDRKG 660  
DB 601 CHTLLCKVOLLSQSESECPDSVQORDVLSGGRHTRVKRKKVTFLEEVTEYIISGDEDRKG 660  
QY 661 PWEFPARDGCRFOKRIQETEDAIGYCLTFEHRERMFNRLOGTCEKGLNVLKQC 713  
DB 661 PWEFPARDGCRFOKRIQETEDAIGYCLTFEHRERMFNRLOGTCEKGLNVLKQC 713

RESULT 2  
AAB95876  
ID AAB95876 standard; protein; 713 AA.  
XX  
AC AAB95876;  
XX  
DT 26-JUN-2001 (first entry)  
XX  
DE Human protein sequence SEQ ID NO:18965.  
XX  
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.  
XX  
OS Homo sapiens.  
XX  
FN EP1074617-A2.  
XX  
PD 07-FEB-2001.  
XX  
PF 28-JUL-2000; 2000EP-00116126.  
XX  
PR 29-JUL-1999; 99JP-00248036.  
PR 27-AUG-1999; 99JP-00300253.  
PR 11-JAN-2000; 2000JP-00118776.  
PR 02-MAY-2000; 2000JP-00183767.  
PR 09-JUN-2000; 2000JP-00241899.  
XX  
PA (HELI-) HELIX RES INST.  
XX  
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX  
DR WPI; 2001-318749/34.  
XX  
PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or



PT diagnosis of the abnormality of the proteins encoded by the full-length  
XX cDNAs.  
PS Claim 8; SEQ ID NO 18965; 2537pp + Sequence Listing; English.  
XX  
CC The present invention describes primer sets for synthesizing 5602 full-  
CC length cDNAs defined in the specification. Where a primer set comprises:  
CC (a) an oligo-dT primer and an oligonucleotide complementary to the  
CC complementary strand of a polynucleotide which comprises one of the 5602  
CC nucleotide sequences defined in the specification, where the  
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
CC of an oligonucleotide comprising a sequence complementary to the  
CC complementary strand of a polynucleotide which comprises a 5'-end  
CC sequence and an oligonucleotide comprising a sequence complementary to a  
CC polynucleotide which comprises a 3'-end sequence, where the  
CC oligonucleotide comprises at least 15 nucleotides and the combination of  
CC the 5'-end sequence/3'-end sequence is selected from those defined in the  
CC specification. The primer sets can be used in antisense therapy and in  
CC gene therapy. The primers are useful for synthesizing polynucleotides,  
CC particularly full-length cDNAs. The primers are also useful for the  
CC detection and/or diagnosis of the abnormality of the proteins encoded by  
CC the full-length cDNAs. The primers allow obtaining of the full-length  
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAH95893  
CC represent human amino acid sequences; and AAH13629 to AAH13632 represent  
CC oligonucleotides, all of which are used in the exemplification of the  
CC present invention  
XX  
SQ Sequence 713 AA;

Query Match 99.7%; Score 3794; DB 4; Length 713;  
Best Local Similarity 99.7%; Pred. No. 5e-302;  
Matches 711; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 MEGTGSRRKLGPRAGFRWPPPPRRSQAGSKKFTPLGPNSSGPTLLSSAQPTRV 60  
DB 1 MEGTGSRRKLGPRAGFRWPPPPRRSQAGSKKFTPLGPNSSGPTLLSSAQPTRV 60  
  
QY 61 SYWTKLSQLLAPLPGLLQKVLQVLSQFGGFFTRWLDFAQVYSGALRALKGRKPAAPTA 120  
DB 61 SYWTKLSQLLAPLPGLLQKVLQVLSQFGGFFTRWLDFAQVYSGALRALKGRKPAAPTA 120  
  
QY 121 QKSLSSQLQSDSDPSVTPDLWLBEGIHQVSPDPLKLELKAGSALDPAQAFLSQOL 180  
DB 121 QKSLSSQLQSDSDPSVTPDLWLBEGIHQVSPDPLKLELKAGSALDPAQAFLSQOL 180  
  
QY 181 WGVLLPSSLOSRLYSNRELSSPSGLNLTORINFSVVLNPSYLDLCPPLREVSQN 240  
DB 181 WGVLLPSSLOSRLYSNRELSSPSGLNLTORINFSVVLNPSYLDLCPPLREVSQN 240  
  
QY 241 SDGNSEVVGFOTLTPSSCLREDHCHPOPLSAELIPASWQCPLSTEGLEPIHHLRMKR 300  
DB 241 SDGNSEVVGFOTLTPSSCLREDHCHPOPLSAELIPASWQCPLSTEGLEPIHHLRMKR 300  
  
QY 301 LEFLQANKGQDLPTPDQNGYHSLBEEHSLRLMDPKHCRDNPTQFVPAAGDIPGNTQES 360  
DB 301 LEFLQANKGQDLPTPDQNGYHSLBEEHSLRLMDPKHCRDNPTQFVPAAGDIPGNTQES 360  
  
QY 361 TEKIELLTITVPLALEESESSECPSEIPEWKEPGEGRISVVVDYSLVLEGLDPLSARPA 420  
DB 361 TEKIELLTITVPLALEESESSECPSEIPEWKEPGEGRISVVVDYSLVLEGLDPLSARPA 420  
  
QY 421 CSNKLIDYILGASDLETSSDPGEDWDEAARDGFDSSLSLSDLEQDPEGLHLWNS 480  
DB 421 CSNKLIDYILGASDLETSSDPGEDWDEAARDGFDSSLSLSDLEQDPEGLHLWNS 480  
  
QY 481 FCSVDVPYNPQNTATIQTAARIVPEEPSDSSEKLSGKSDLENSSQSGSLPETHSGEE 540  
DB 481 FCSVDVPYNPQNTATIQTAARIVPEEPSDSSEKLSGKSDLENSSQSGSLPETHSGEE 540  
  
QY 541 DDWESSADEAESKLWNSFCNSDDPYNPLNFKAPQTSGENEKCRCRSTPESIVASE 600  
DB 541 DDWESSADEAESKLWNSFCNSDDPYNPLNFKAPQTSGENEKCRCRSTPESIVASE 600

QY 601 CHTLLSCKVQLLSQSESECPDSVQRDVLSGGRHTRHVKRKKVTFLEEVTEYIISGDEDRKG 660  
DB 601 CHTLLSCKVQLLSQSESECPDSVQRDVLSGGRHTRHVKRKKVTFLEEVTEYIISGDEDRKG 660  
  
QY 661 PWEEFARDGCRFQKRIQETEDATGYCLTTEHRRMFRNRLQGTCTFKGLNVLKQC 713  
DB 661 PWEEFARDGCRFQKRIQETEDATGYCLTTEHRRMFRNRLQGTCTFKGLNVLKQC 713  
  
RESULT 3  
AAB53401  
ID AAB53401 standard; protein; 707 AA.  
XX AAB53401;  
AC AAB53401;  
XX  
DT 09-MAR-2001 (first entry)  
DE Human colon cancer antigen protein sequence SEQ ID NO:941.  
XX  
KW Human; colon cancer; colon cancer antigen; diagnosis; detection;  
KW identification; cytostatic; cardioactive; neuroprotective; vulnary;  
KW immunomodulatory; muscular; gynaecological; gastrointestinal;  
KW nephrotropic; antiinfective; antibacterial; gene therapy; wound;  
KW neural disorder; immune system disorder; muscular disorder;  
KW reproductive disorder; gastrointestinal disorder; renal disorder;  
KW infectious disease; cardiovascular disorder.  
XX  
OS Homo sapiens.  
XX  
PN W0200055351-A1.  
XX  
PD 21-SEP-2000.  
XX  
PF 08-MAR-2000; 2000MO-US005883.  
XX  
PR 12-MAR-1999; 99US-0124270P.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Rosen CA, Ruben SM;  
XX  
DR WPI; 2000-587534/55.  
XX  
DR N-PSDB; AAC98158.  
XX  
PT Colon cancer associated gene sequences, referred to as colon cancer  
PT antigens, useful for the treatment, prevention, and diagnosis of colon  
PT disorders such as colon cancer.  
XX  
PS Claim 11; Page 1503-1506; 2104pp; English.  
XX  
CC AAC97991 to AAC98763 encode the human colon cancer associated proteins,  
CC called human colon cancer antigens, given in AAB53234 to AAB54006. The  
CC human colon cancer antigens can have cytostatic, cardioactive, muscular;  
CC neuroprotective, immunomodulatory, gynaecological, gastrointestinal, and  
CC vulnary, nephrotropic, antiinfective and antibacterial activities, and  
CC can be used in gene therapy. The colon cancer antigen polynucleotides,  
CC proteins and antibodies to the proteins are useful for the prevention,  
CC treatment and diagnosis of colon disorders, such as colon cancer. The  
CC polynucleotides may be used in diagnostics and research, such as for  
CC chromosome identification, and as hybridisation probes. The proteins may  
CC also be used to prevent diseases such as neural disorders, immune system  
CC disorders, muscular disorders, reproductive disorders, gastrointestinal  
CC disorders, wounds, renal disorders, infectious diseases, and  
CC cardiovascular disorders. AAC98764 to AAC98772 and AAB54007 represent  
CC sequences used in the exemplification of the present invention  
XX  
SQ Sequence 707 AA;

Query Match 95.3%; Score 3627; DB 3; Length 707;  
Best Local Similarity 98.8%; Pred. No. 2.5e-288;  
Matches 683; Conservative 1; Mismatches 7; Indels 0; Gaps 0;







[illegible]

RESIN.T 8

RESULI 8  
ADD45119  
ID ADD45119 standard: protein: 674 AA

AC ADD45119:

29-JAN-2004 (first entry)

XX  
XX  
(УТННН 3АТТ) 4002-МНО-ЕЗ ТД

DE Human Protein XP 009097, SEQ ID NO 10552.

22 XX  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
841  
842  
843  
844  
845  
846  
847  
848  
849  
850  
851  
852  
853  
854  
855  
8

Human; pain; neuronal tissue; gene therapy;

spinal segmental nerve injury; chronic constriction injury; CCI; KW

KW spared nerve injury; SNI; Chung.

XX  
XX

OS Homo sapiens.

4

PN WO2003016475-A2.

XX

PD 27-FEB-2003.

XX

PF 14-AUG-2002; 2002WO-US025765.

6  
 7  
 8  
 9  
 10  
 11  
 12  
 13  
 14  
 15  
 16  
 17  
 18  
 19  
 20  
 21  
 22  
 23  
 24  
 25  
 26  
 27  
 28  
 29  
 30  
 31  
 32  
 33  
 34  
 35  
 36  
 37  
 38  
 39  
 40  
 41  
 42  
 43  
 44  
 45  
 46  
 47  
 48  
 49  
 50  
 51  
 52  
 53  
 54  
 55  
 56  
 57  
 58  
 59  
 60  
 61  
 62  
 63  
 64  
 65  
 66  
 67  
 68  
 69  
 70  
 71  
 72  
 73  
 74  
 75  
 76  
 77  
 78  
 79  
 80  
 81  
 82  
 83  
 84  
 85  
 86  
 87  
 88  
 89  
 90  
 91  
 92  
 93  
 94  
 95  
 96  
 97  
 98  
 99  
 100  
 101  
 102  
 103  
 104  
 105  
 106  
 107  
 108  
 109  
 110  
 111  
 112  
 113  
 114  
 115  
 116  
 117  
 118  
 119  
 120  
 121  
 122  
 123  
 124  
 125  
 126  
 127  
 128  
 129  
 130  
 131  
 132  
 133  
 134  
 135  
 136  
 137  
 138  
 139  
 140  
 141  
 142  
 143  
 144  
 145  
 146  
 147  
 148  
 149  
 150  
 151  
 152  
 153  
 154  
 155  
 156  
 157  
 158  
 159  
 160  
 161  
 162  
 163  
 164  
 165  
 166  
 167  
 168  
 169  
 170  
 171  
 172  
 173  
 174  
 175  
 176  
 177  
 178  
 179  
 180  
 181  
 182  
 183  
 184  
 185  
 186  
 187  
 188  
 189  
 190  
 191  
 192  
 193  
 194  
 195  
 196  
 197  
 198  
 199  
 200  
 201  
 202  
 203  
 204  
 205  
 206  
 207  
 208  
 209  
 210  
 211  
 212  
 213  
 214  
 215  
 216  
 217  
 218  
 219  
 220  
 221  
 222  
 223  
 224  
 225  
 226  
 227  
 228  
 229  
 230  
 231  
 232  
 233  
 234  
 235  
 236  
 237  
 238  
 239  
 240  
 241  
 242  
 243  
 244  
 245  
 246  
 247  
 248  
 249  
 250  
 251  
 252  
 253  
 254  
 255  
 256  
 257  
 258  
 259  
 260  
 261  
 262  
 263  
 264  
 265  
 266  
 267  
 268  
 269  
 270  
 271  
 272  
 273  
 274  
 275  
 276  
 277  
 278  
 279  
 280  
 281  
 282  
 283  
 284  
 285  
 286  
 287  
 288  
 289  
 290  
 291  
 292  
 293  
 294  
 295  
 296  
 297  
 298  
 299  
 300  
 301  
 302  
 303  
 304  
 305  
 306  
 307  
 308  
 309  
 310  
 311  
 312  
 313  
 314  
 315  
 316  
 317  
 318  
 319  
 320  
 321  
 322  
 323  
 324  
 325  
 326  
 327  
 328  
 329  
 330  
 331  
 332  
 333  
 334  
 335  
 336  
 337  
 338  
 339  
 340  
 341  
 342  
 343  
 344  
 345  
 346  
 347  
 348  
 349  
 350  
 351  
 352  
 353  
 354  
 355  
 356  
 357  
 358  
 359  
 360  
 361  
 362  
 363  
 364  
 365  
 366  
 367  
 368  
 369  
 370  
 371  
 372  
 373  
 374  
 375  
 376  
 377  
 378  
 379  
 380  
 381  
 382  
 383  
 384  
 385  
 386  
 387  
 388  
 389  
 390  
 391  
 392  
 393  
 394  
 395  
 396  
 397  
 398  
 399  
 400  
 401  
 402  
 403  
 404  
 405  
 406  
 407  
 408  
 409  
 410  
 411  
 412  
 413  
 414  
 415  
 416  
 417  
 418  
 419  
 420  
 421  
 422  
 423  
 424  
 425  
 426  
 427  
 428  
 429  
 430  
 431  
 432  
 433  
 434  
 435  
 436  
 437  
 438  
 439  
 440  
 441  
 442  
 443  
 444  
 445  
 446  
 447  
 448  
 449  
 450  
 451  
 452  
 453  
 454  
 455  
 456  
 457  
 458  
 459  
 460  
 461  
 462  
 463  
 464  
 465  
 466  
 467  
 468  
 469  
 470  
 471  
 472  
 473  
 474  
 475  
 476  
 477  
 478  
 479  
 480  
 481  
 482  
 483  
 484  
 485  
 486  
 487  
 488  
 489  
 490  
 491  
 492  
 493  
 494  
 495  
 496  
 497  
 498  
 499  
 500  
 501  
 502  
 503  
 504  
 505  
 506  
 507  
 508  
 509  
 510  
 511  
 512  
 513  
 514  
 515  
 516  
 517  
 518  
 519  
 520  
 521  
 522  
 523  
 524  
 525  
 526  
 527  
 528  
 529

PR 14-AUG-2001; 2001US-0312147P.

PR 01-NOV-2001; 2001US-0346382P.  
PB 26-NOV-2001; 2001US-0323247P.  
PB 01-NOV-2001; 2001US-0323247P.

PR 26-NOV-2001; 2001US-0333334  
yy

XX  
PA (GENO) GEN HOSPT

PA (GEHU) GEN HOSPITAL CORP.  
PA (FAPB) BAYER AG  
PA (FAPB) BAYER AG

PA (FARB) BAYER AG.  
XX

PT Woolf C. D'urso D. Refort K. Costigan M.

FI MOULT C, D'ALBO D, BERTOLIC K, CORVETTUAN M, XX

WIPI; 2003-268312/26.  
 GENBANK; XP\_009097.  
 New composition comprising two or more isolated polypeptides, useful for  
 preparing a medicament for treating pain in an animal.  
 Claim 1; Page; 1017pp; English.  
 The invention discloses a composition comprising two or more isolated rat  
 or human polynucleotides or a polynucleotide which represents a fragment  
 derivative or allelic variation of the nucleic acid sequence. Also  
 claimed are a vector comprising the novel polynucleotide, a host cell  
 comprising the vector, a method for identifying a nucleotide sequence  
 which is differentially regulated in an animal subjected to pain and a  
 kit to perform the method, an array, a method for identifying an agent  
 that increases or decreases the expression of the polynucleotide sequence  
 that is differentially expressed in neuronal tissue of a first animal  
 subjected to pain, a method for identifying a compound which regulates  
 the expression of a polynucleotide sequence which is differentially  
 expressed in an animal subjected to pain, a method for identifying a  
 compound that regulates the activity of one or more of the  
 polynucleotides, a method for producing a pharmaceutical composition, a  
 method for identifying a compound or small molecule that regulates the  
 activity in an animal of one or more of the polypeptides given in the  
 specification, a method for identifying a compound useful in treating  
 pain and a pharmaceutical composition comprising the one or more  
 polypeptides or their antibodies. The polynucleotide or the compound that  
 modulates its activity is useful for preparing a medicament for treating  
 pain (e.g. spinal segmental nerve injury (Chung), chronic constriction  
 injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene  
 therapy). The sequence presented is a human protein (shown in Table 2 of  
 the specification) which is differentially expressed during pain. Note:  
 The sequence data for this patent did not form part of the printed  
 specification, but was obtained in electronic form directly from WIPO at  
 ftp.wipo.int/pub/published\_pct\_sequences.  
 Sequence 674 AA:

Query Match 6.8%; Score 260.5; DB 7; Length 674;  
Best Local Similarity 21.9%; Pred. No. 6.3e-12;  
Matches 164; Conservative 80; Mismatches 263; Indels 241; Gaps 32

[illegible]









[illegible]



Db 199 STVPFLGEA-----EQATEEK-----GTE-NKADPSNPSGSGSHRAWEYYSREKPKQ 247  
 Qy 399 GRISVVDVSYLGDLPI SARPA-----CSNKILIDVILGASDLTSDP-----EG 445  
 Db 248 -----EGEAKVEAHRAQGHPCRNABAE--GGPETTFVCTGNFLKAWVYRPG 294  
 Qy 446 EDWDEE-----AEDD-----GFSDSLSLSD-LEQDPEGL- 475  
 Db 295 EDTEEDNSDSAEEDTAQTGATHTSAFLKAWVYRPGEDTEEDSDSAEEDTAGT 354  
 Qy 476 ---HLWNSFCSDVPYNPQNT-----ATIQTAARI---VPEEPPDSE 511  
 Db 355 ATPHT-SAFKAWVYRPGEDTEEDNSDSAEEDTAQTGATHTSAFLKAWVYRPGEDTE 413  
 Qy 512 KDLSCKSLD-----ENSSQSGSLPET-----PEHSSGEEDWES-----S 546  
 Db 414 EB-----NSDLSAEEDTAQTGATHTSPFLKAWVYRPGEDTEEDTEEDSENAPGDSET 470  
 Qy 547 ADEAESLKLWNSFC-----NSDDPYNPLNFKAPFQTSGENEKGCRDSTKTPSESIVAI 598  
 Db 471 ADSSQSPCLQPCORCLPGEKTKRGEEP--PL-FQVAFYLPGEK-----PESPWAA 517  
 Qy 599 SECHTLCKVQLGS-----QSECPDSVQRDLVLSGGRHTHVKKVTFLEEVTEYIS- 653  
 Db 518 PKLPLRLQRLRLFKAPTRDQDPEIP-----LKARKVHFAEKTVMHFLAV 562  
 Qy 654 ----GDEDRKGWEFARDCGCFQKRIQETEDAIGYCLTFEHRERFNRQ 700  
 Db 563 WAGPAQAARRGPEQFARDRSRFAARIAQAEEKLGPYLTTPDSRARAWARL 613

RESULT 14

ABBS7373  
 ID ABBS7373 standard; protein; 657 AA.

AC ABBS7373;

DT 07-MAR-2002 (first entry)

DE Mouse ischaemic condition related protein sequence SEQ ID NO:1062.

KW Mouse; ischaemia; compressive ischaemia; occlusive ischaemia;  
 KW vasospastic ischaemia; ischaemic condition; ischaemic disease.

OS Mus musculus.

PN WO200188188-A2.

XX 22-NOV-2001.

XX 18-MAY-2001; 2001WO-JP004192.

XX 18-MAY-2000; 2000JP-00145977.

PA (UTNI-) UNIV NIHON SCHOOL JURIDICAL PERSON.

PI Ishikawa K, Asai S, Takahashi Y, Nagata T, Ishii Y;

DR WPI; 2002-034733/04.

DR N-PSDB; ABI99909.

XX Examining the ischaemic condition (e.g. occlusive ischaemia) by measuring  
 XX expression levels of particular genes defined in the specification or by  
 XX determining the expression profile of a gene group comprising these  
 XX genes.

PS Claim 2; Page 2676-2679; 2690pp; English.

XX The present invention describes a method for examining ischaemic  
 XX conditions, comprising measuring the expression levels of particular  
 XX genes (I) in a test sample or determining the expression profile of a  
 XX gene group in the sample comprising genes selected from (I). The method

CC is useful for examining the ischaemic condition (e.g. compressive  
 CC ischaemia, occlusive ischaemia or vasospastic ischaemia) by measuring  
 CC expression levels of particular genes (ABI99909 to ABI9912, encoding the  
 CC protein sequences in ABBS7020 to ABBS7374) or by determining the  
 CC expression profile of a gene group comprising these genes. The expression  
 CC levels or expression profiles produced by these genes are used as an  
 CC indicator when screening for ischaemic condition-improving drugs or  
 CC therapeutics for ischaemic diseases. ABI99913 and ABI99914 represent PCR  
 CC primers for a mouse ischaemic condition related sequence, which are used  
 CC in the exemplification of the present invention

XX SQ Sequence 657 AA;

Query Match 6.2%; Score 237; DB 5; Length 657;  
 Best Local Similarity 21.8%; Pred. No. 5.2e-10;  
 Matches 169; Conservative 81; Mismatches 233; Indels 292; Gaps 35;  
 Qy 53 SAQPETRVSVYTKLLS-QLLAPLPLGLLQKLVILMSQLFGMPTRWLDFAGVYSALRAKLG 111  
 Db 4 SPRFQ-HVLHWRDAHNFYLLSPLMGLLSRA--WSRLRGPEVPBWL----- 46  
 Qy 112 REXPAAPTAQKSLSLSDSDSPSVTPSLDWELEGTHWQYSPDLKLEKAKGSALDPAA 171  
 Db 47 -----AKVTGADOIE----- 58  
 Qy 172 QAFLEQQLMGVELLPSSLSQSLRYSNRELSSPSGSLNIOKIDNFSVVYLLNPSYLDGF 231  
 Db 59 AALLTPFVSG-NLLP-----HGETEESGSPESQAAQRL-----CL 94  
 Qy 232 PRLEVSQNSDGNSEVVGFTLTPESCLREDHCHPOPLSAELIPASWQCPPLSTGLP 291  
 Db 95 VEAESSPPETWGLSNVDEYNA-KPGQDDLREKEMERTAGKATLQAPAGLOAD----- 145  
 Qy 292 EIHHLRMKRLFLQOANKG---QDLTPDQDNGVHSLSEESLLR-----MDPKHCR 340  
 Db 146 -----KRLGEVVAREEGVAEPAYPTSQLEGGPAENEEDGETVTKTVOAASATAPYKP 198  
 Qy 341 DNPTQFVPAAGDIPGNTQESTEEKIELLTTEVPLALAEESPSGEGCPSS--EIPWEKEPGE 398  
 Db 199 STVPFLGEA-----EQATEEK-----GTE-NKADPSNPSGSGSHRAWEYYSREKPKQ 247  
 Qy 399 GRISVVDVSYLGDLPI SARPA-----CSNKILID-----YILGAS 434  
 Db 248 -----EGEAKVEAHRAQGHPCRNABAE--GGPETTFVCTGNFLKAWVYRPGED 296  
 Qy 435 SDLETSSDPE-----GEDWDERAEDDGFSDSLSLSD 468  
 Db 297 TEEDNSDSAEEDTAQTGATHTSAFLKAWVYRPGEDTEER-----DSDSDSAEDT 350  
 Qy 469 EQDPEGLHLWNSFCSDVPYNPQNT-----ATIQTAARI---VPEEP 507  
 Db 351 AQTGATPHT-SAFKAWVYRPGEDTEEDNSDSAEEDTAQTGATHTSAFLKAWVYRPG 409  
 Qy 508 SDSEKDLGKSDL-----ENSSQSGSLPET-----PEHSSGEEDWES----- 545  
 Db 410 EDTEEE-----NSDLSAEEDTAQTGATHTSPFLKAWVYRPGEDTEEDTEEDSENAPG 466  
 Qy 546 ---SADEAESLKLWNSFC-----NSDDPYNPLNFKAPFQTSGENEKGCRDSTKTPSES 594  
 Db 467 DSETADSSQSPCLQPCORCLPGEKTKRGEEP--PL-FQVAFYLPGEK-----PES 513  
 Qy 595 IVAISECHTLCKVQLGS-----QSECPDSVQRDLVLSGGRHTHVKKVTFLEEVTEY 650  
 Db 514 PWAAPKLPLRLQRLRLFKAPTRDQDPEIP-----LKARKVHFAEKTVMHFLAV 558  
 Qy 651 YIS-----GDEDRKGWEFARDCGCFQKRIQETEDAIGYCLTFEHRERFNRQ 700  
 Db 559 FLAVWAGPAQAARRGPEQFARDRSRFAARIAQAEEKLGPYLTTPDSRARAWARL 613

RESULT 15

AAW79958  
 ID AAW79958 standard; protein; 578 AA.



**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 16, 2005, 09:17:12 ; Search time 19.202 Seconds  
(without alignments)  
2771.838 Million cell updates/sec

Title: US-10-650-482-2

Perfect score: 3804

Sequence: 1 MEFPTGGSRKRLGPRAGFRP.....RMFNRLQGTCTGKGLNLVKQC 713

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : Issued Patents\_AA.\*

1: /cgn2\_6/ptodata/1/1aa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/1/1aa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/1/1aa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/1/1aa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/1/1aa/PCTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/1/1aa/backfiles.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	252.5	6.6	674	3	US-08-893-852A-1
2	237	6.2	657	3	US-08-893-852A-3
3	236	6.2	657	3	US-08-821-818-3
4	236	6.2	657	4	US-09-052-753B-3
5	219	5.8	578	4	US-09-052-753B-7
6	201	5.3	590	3	US-08-893-852A-4
7	201	5.3	590	3	US-08-821-818-2
8	201	5.3	590	4	US-09-052-753B-2
9	167	4.4	1596	4	US-08-978-277A-4
10	156.5	4.1	1503	3	US-08-976-255-14
11	148	3.9	1142	2	US-08-993-118-7
12	148	3.9	1142	3	US-08-845-528C-7
13	148	3.9	1142	4	US-09-066-281B-7
14	148	3.9	1142	4	US-09-468-433C-7
15	141	3.7	993	3	US-09-060-410-4
16	141	3.7	993	4	US-09-723-458-4
17	139	3.7	823	4	US-09-248-796A-16699
18	139	3.7	1346	2	US-08-635-121-2
19	139	3.7	1346	4	US-08-978-277A-2
20	138	3.6	1142	3	US-09-061-709-2
21	138	3.6	1142	4	US-09-899-651-2
22	138	3.6	1142	4	US-09-392-714-26
23	138	3.6	1142	4	US-09-270-437D-2
24	137	3.6	901	4	US-09-248-796A-14747
25	135.5	3.6	581	4	US-09-244-805-27
26	135	3.5	414	4	US-09-248-796A-19046
27	134.5	3.5	773	4	US-09-270-767-33141

28	134.5	3.5	773	4	US-09-270-767-48358	Sequence 48358, A
29	134.5	3.5	1259	4	US-09-949-016-10366	Sequence 10366, A
30	134.5	3.5	1969	4	US-09-418-710-72	Sequence 72, Appl
31	134.5	3.5	1969	4	US-09-839-479-71	Sequence 71, Appl
32	134.5	3.5	2468	4	US-09-976-594-726	Sequence 726, Appl
33	134.5	3.5	2468	4	US-09-538-092-1135	Sequence 1135, Ap
34	134.5	3.5	2522	4	US-09-949-016-10337	Sequence 10337, A
35	133	3.5	513	4	US-09-949-016-6492	Sequence 6492, Ap
36	133	3.5	534	4	US-09-949-016-10364	Sequence 10364, A
37	133	3.5	933	3	US-08-293-728-2	Sequence 2, Appl1
38	133	3.5	933	3	US-09-421-868-2	Sequence 2, Appl1
39	133	3.5	1200	4	US-09-644-827B-8	Sequence 8, Appl1
40	133	3.5	1428	4	US-09-644-827B-7	Sequence 7, Appl1
41	132.5	3.5	1972	4	US-09-418-710-21	Sequence 21, Appl
42	132.5	3.5	1972	4	US-09-839-479-21	Sequence 21, Appl
43	132	3.5	936	4	US-08-956-171E-5249	Sequence 5249, Ap
44	132	3.5	936	4	US-08-781-986A-5249	Sequence 5249, Ap
45	132	3.5	2753	4	US-09-949-016-7659	Sequence 7659, Ap
46	132	3.5	2753	4	US-09-949-016-7660	Sequence 7660, Ap
47	131.5	3.5	1435	2	US-08-568-459A-4	Sequence 4, Appl1
48	131.5	3.5	1435	2	US-08-487-826B-4	Sequence 4, Appl1
49	131.5	3.5	1435	3	US-09-210-288-4	Sequence 4, Appl1
50	131	3.4	693	4	US-09-081-385-154	Sequence 154, App
51	131	3.4	879	4	US-09-248-796A-20328	Sequence 20328, A
52	131	3.4	937	4	US-09-447-399-4	Sequence 4, Appl1
53	131	3.4	1048	3	US-09-356-952-5	Sequence 5, Appl1
54	130.5	3.4	377	4	US-09-248-796A-20227	Sequence 20227, A
55	130.5	3.4	808	4	US-09-270-767-36557	Sequence 36557, A
56	130.5	3.4	808	4	US-09-270-767-51774	Sequence 51774, A
57	130	3.4	781	4	US-09-949-016-9773	Sequence 9773, Ap
58	130	3.4	1805	1	US-07-853-913-2	Sequence 2, Appl1
59	129	3.4	764	4	US-09-538-092-944	Sequence 944, App
60	129	3.4	1664	1	US-09-599-652-2	Sequence 2, Appl1
61	129	3.4	1664	2	US-08-642-846-2	Sequence 2, Appl1
62	129	3.4	1664	3	US-09-264-604-2	Sequence 2, Appl1
63	129	3.4	1664	4	US-09-378-343-2	Sequence 2, Appl1
64	128	3.4	927	4	US-09-248-796A-14284	Sequence 14284, A
65	128	3.4	1087	1	US-08-264-002-5	Sequence 5, Appl1
66	128	3.4	1780	1	US-08-769-309A-5	Sequence 5, Appl1
67	128	3.4	1780	3	US-08-994-570-5	Sequence 5, Appl1
68	128	3.4	1781	4	US-09-961-403-13	Sequence 13, Appl
69	128	3.4	3924	4	US-09-538-092-1246	Sequence 1246, Ap
70	127.5	3.4	568	1	US-08-320-559-30	Sequence 30, Appl
71	127.5	3.4	568	3	US-08-545-860D-30	Sequence 30, Appl
72	127.5	3.4	568	4	US-09-538-092-1114	Sequence 1114, Ap
73	127.5	3.4	568	5	PCT-US94-04496-30	Sequence 30, Appl
74	127	3.3	557	4	US-09-248-796A-19073	Sequence 19073, A
75	127	3.3	784	4	US-09-370-838-67	Sequence 67, Appl
76	127	3.3	784	4	US-09-854-133-67	Sequence 67, Appl
77	127	3.3	783	6	5231168-2	Patent No. 5231168
78	127	3.3	783	6	5231168-2	Patent No. 5231168
79	127	3.3	1317	3	US-09-083-521-7	Sequence 7, Appl1
80	126.5	3.3	2476	3	US-09-824-574-7	Sequence 7, Appl1
81	125.5	3.3	1257	1	US-08-340-428B-49	Sequence 49, Appl
82	124.5	3.3	473	4	US-09-248-796A-27126	Sequence 27126, A
83	124	3.3	830	4	US-09-562-737-39	Sequence 39, Appl
84	124	3.3	978	4	US-09-949-016-7281	Sequence 7281, Ap
85	124	3.3	3256	4	US-09-919-172-98	Sequence 98, Appl
86	124	3.3	3256	4	US-09-976-594-22	Sequence 22, Appl
87	124	3.3	3256	4	US-09-319-039-21	Sequence 21, Appl
88	124	3.3	3913	4	US-09-949-016-10933	Sequence 10933, A
89	124	3.3	4377	4	US-09-949-016-6978	Sequence 6978, Ap
90	123.5	3.2	1166	4	US-09-200-650B-7	Sequence 7, Appl1
91	123.5	3.2	1878	4	US-09-949-016-8902	Sequence 8902, Ap
92	123.5	3.2	1878	4	US-09-248-796A-21836	Sequence 21836, A
93	123	3.2	385	4	US-08-317-522A-5	Sequence 5, Appl1
94	123	3.2	732	1	US-09-562-737-33	Sequence 33, Appl
95	123	3.2	830	4	US-09-200-650B-3	Sequence 3, Appl1
96	123	3.2	930	4	US-09-344-624-4	Sequence 4, Appl1
97	123	3.2	1219	4	US-09-200-650B-1	Sequence 1, Appl1
98	122.5	3.2	918	4	US-08-320-559-28	Sequence 28, Appl
99	122.5	3.2	1187	1	US-08-545-860D-28	Sequence 28, Appl
100	122.5	3.2	1187	3		

## ALIGNMENTS

```

RESULT 1
US-08-893-852A-1
; Sequence 1, Application US/08893852A
; Patent No. 6080558
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Lal, Preeti
; APPLICANT: Shah, Purvi
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: NEW HUMAN GROWTH REGULATOR PROTEIN
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESS: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/893,852A
; FILING DATE: Filed Herewith
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PP-0341 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 674 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: TMLR3DT01
; CLONE: 508302
US-08-893-852A-1

Query Match 6.6%; Score 252.5; DB 3; Length 674;
Best Local Similarity 21.9%; Pred. No. 9.1e-13;
Matches 164; Conservative 80; Mismatches 262; Indels 243; Gaps 32;

QY 70 LLAPLGLQKVLINSLQGLFGMFPTR-WLDAGVYSAL--RALKGR-----112
DB 21 LLSPVMSLLSRA--WSRL-RGLGPLEPWLVEAVKGAALVEAGLEARTPLAIPHTPWGR 77
QY 113 -----EKPAAPTAQKSLSSQLDSDPSVTSPDLWLEEGIHWOYSPDPLKLELKAGSA 166
DB 78 RPEEAEDSGGDETRTLGLTKTSSSLPEAWGLLD-DDDGMVGERATSV--RGQSQ 133
QY 167 LDPAQAFLLEQQLMGVLLPSSLRYSNRELGSFSGPLNIQRIIDNFVSVLWPS 226
DB 134 FADGORAPL-----SPSLIIRTLQSDKNPGEAKAEEGVAAEEGVNKF-----YPPS 182
QY 227 YLDCPRLEVSQNSDGNSEVVGFTLTPESCLREDDH-CHPQPLSAELIPASWGGCPPL 285
DB 183 HRECCPAVE---EEDD-----EAAVKKAHRTSTLSALPGSKFSTWVSCB-- 224
QY 286 STEGLPEIHLMKRLEFLQANKQDLFTPDQDNGVHSLBEEHSLLRMDPKHCRDNPQTQ 345

```

```

DB 225 ---GEENQATEDKRTERSKGARK-----TSVSPRSGSDPRS 259
QY 346 FVPAAGDIPGNTQESTEEKIELLITTEVPLALEEESPSEGCPSSEI PKEKPGEGRISVVD 405
DB 260 WEYRSG-----EASEKEE-----KAHEETGKGEAAPG-----287
QY 406 YSYLEGLDLPISARPACSNKLIDY-----ILGASSDLETSSDP-----443
DB 288 -----POSSAPAQRPQLKSWWCQPSDRESEVKALGAEKDGEAECPCIPPPSAFL 339
QY 444 -----EGEDWDEEAEDGFDSDSLSDLEQDPEG-----LHLWNSFCSDPY 487
DB 340 KAWYWPGEDETEE-EDDEEDSDSGSDEEAGEAASSTPATGVFLKSW-----VY 391
QY 488 NPQNFTATIQTAAARIVPEEPSDEKDLGSKSLDLENSQSGSLP-----ET 532
DB 392 QPGEDT-----EEDDEDSDTGSADEREAEATSASTPPASAFKAWYVRPGEDT 440
QY 533 PEH-----SSGEEDDWESSADEAESLKLWNSFCNSDDPYNPLNFKAPFOTSGENKGCGR 586
DB 441 EEEDEDVDSKEDDSEALGEAES-----DHP-----SHPDQSAHFRGWGR 485
QY 587 DSKTPSSIVA-----ISECHTLCLKVQLLGSQ-----ESECPSDVQDVLGGRH 635
DB 486 PGKETEEREAADWGEAEPCCFRVA--IYVPGKPPPPWAPPRLPLRLQRLKRPETPTH 543
QY 636 -----VKRKKYTFLEEVTEYYIS-----GDEDRKGPWEFEFARDGCRFOKRIQETDAI 683
DB 544 DDPETPLKARKVRFSEKVTVHFLAVWAGPAARQGWESQLARDRSRRFARRITQAEEL 603
QY 684 GYCLTFEHRERMFNRLOGTCFKGLNVLKQ 712
DB 604 SPCLTPAARARAWARLENPPLAIPALTO 632

```

## RESULT 2

```

US-08-893-852A-3
; Sequence 3, Application US/08893852A
; Patent No. 6080558
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Lal, Preeti
; APPLICANT: Shah, Purvi
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: NEW HUMAN GROWTH REGULATOR PROTEIN
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESS: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/893,852A
; FILING DATE: Filed Herewith
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PP-0341 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 3:

```



SEQUENCE CHARACTERISTICS:  
 LENGTH: 657 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 IMMEDIATE SOURCE:  
 LIBRARY: GenBank  
 CLONE: 53041  
 US-08-893-852A-3

Query Match 6.2%; Score 237; DB 3; Length 657;  
 Best Local Similarity 21.8%; Pred. No. 1.9e-11;  
 Matches 169; Conservative 81; Mismatches 233; Indels 292; Gaps 35;  
 QY 53 SAQPTETRVSYTKLLS-QLLAPLQGLKVLKQVLSQFGMPPTWLDPAQVYSALRALKG 111  
 DB 4 SPRFQ-HVLRWRDHNHFLVLSPLMGLLSRA--WSRLRGPEVPEAWL----- 46  
 QY 112 REKPAAPTAQKSLSLQLDSSDPVSTPDLWLEEGIHQVYSPDLKLELKAKGSALDPAA 171  
 DB 47 -----AKTVTGADQIE----- 58  
 QY 172 QAFLEQOLKGVLLPSLQRLVSNRELSSPGPLNIQRIDNFSVYLLNPSYLDGP 231  
 DB 59 AALLTPTPVSG-NLLP-----HGETEESGSPQSQAQRL-----CL 94  
 QY 232 PRLVSVYQNSDGNSEVVGFTLTPESSCLREDHCHPQLSAELIPASWQGPCPLSTEGLP 291  
 DB 95 VEAESSPPETWGLNVDIYNA-KGQDRLREKEMERTAGKATLQAPLQAGAD----- 145  
 QY 292 EIHRLMKRLEFLOQANKG---QDLPTPDQNGYHSLSEHSLLR-----MDPKHCR 340  
 DB 146 -----KRLGEVAREEGVAEPVPTSQLGGPAENEDEGETVKTYQASAASTAPGKYP 198  
 QY 341 DNPTQFPAAGDIPGNTQESTTEKIELTTVPVLALEESSEGCPS--EIPMEKEPGE 398  
 DB 199 STPVVFLGEA-----EHQATEEK---GTE-NKADPSNPSGSHSRAWEYYSREKPKQ 247  
 QY 399 GRISVDVSYLEGDLPI SARPA-----CSNKLID-----YILGAS 434  
 DB 248 -----EGEAKVEAHRAGQGHPCRNABEEGGPPTFVCTGNALFKAWYRPGED 296  
 QY 435 SDLETSSDPE-----GEDWDEAEEDGDFDSDSLSDSL 468  
 DB 297 TEEDNSDSDAEEDTAQTGATPHTSAFLKAWYRPGEDTEEE-----DSDSDAEEDT 350  
 QY 469 EQDPEGLHWNPSVCSVDYNPQNT-----ATIQTAARI-----VPEEP 507  
 DB 351 AQTGATPHT-SAFKAWYRPGEDTEEBNSDLSDAEEDTAQTGATPHTSAFLKAWYRPG 409  
 QY 508 SDSEKOLSGKSDL-----ENSSQSGSLPET-----PEHSSGREDDWES--- 545  
 DB 410 EDTTEE---NSDLSAEEDTAQTGATPHTSAFLKAWYRPGEDTEEDTEEBESENAPG 466  
 QY 546 ---SADAEESLKLWNSFC-----NSDDPYNPLNFKAPFQTSGENEKGCRDSTPSES 594  
 DB 467 DSETADSSQSLQPCRLCPGEKTKGRGEPP-PL-FQVAYFLPGEK-----PES 513  
 QY 595 IVAISECHTLSCVKQLLGS-----QSECPDSVORDVLGGRRHTRVKKKVTFLVEVTEY 650  
 DB 514 PWAAPKPLRLQRLRLFPKAPTRDQDPEIP-----LKARKVHFAEKVTVH 558  
 QY 651 YIS-----GDEDRKGPWEAFARDGCRFQKIQETEDAIGYCLTPEHREMERNLQ 700  
 DB 559 FLAVWAGPAARRGPMWQFARDRSFARRIAQAEEKLGLPYLTPDSRARAWRLR 613

## RESULT 3

US-08-821-818-3  
 ; Sequence 3, Application US/08821818  
 ; Patent No. 6146877  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Fisher, Paul B.

TITLE OF INVENTION: IDENTIFICATION OF THE PROGRESSION  
 TITLE OF INVENTION: ELEVATED GENE-3 AND USES THEREOF  
 NUMBER OF SEQUENCES: 5  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Cooper & Dunham LLP  
 STREET: 1185 Avenue of the Americas  
 CITY: New York  
 STATE: New York  
 COUNTRY: USA  
 ZIP: 10036  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/821,818  
 FILING DATE:  
 CLASSIFICATION: 514  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Chan, Albert Wai-Kit  
 REGISTRATION NUMBER: 36,479  
 REFERENCE/DOCKET NUMBER: 51523  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (212) 278-0400  
 TELEFAX: (212) 391-0525  
 INFORMATION FOR SEQ ID NO: 3:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 657 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-821-818-3

Query Match 6.2%; Score 236; DB 3; Length 657;  
 Best Local Similarity 21.8%; Pred. No. 2.3e-11;  
 Matches 165; Conservative 77; Mismatches 225; Indels 290; Gaps 33;  
 QY 70 LLAPLQGLKVLKQVLSQFGMPPTWLDPAQVYSALRALKREKPAAPTAQKSLSLQL 129  
 DB 21 LLSPLMGLLSRA--WSRLRGPEVPEAWL----- 46  
 QY 130 DSDPSVSTPLDLEEGIHQVYSPDLKLELKAKGSALDPAAQAFLEQOLGWVLLPSS 189  
 DB 47 -----AKTVTGADQIE-----AAALLTPTPVSG-NLLP-- 73  
 QY 190 LQRLVSNRELSSPGPLNIQRIDNFSVYLLNPSYLDPCPRLEVSQNSDGNSEVVG 249  
 DB 74 -----HGETEESGSPQSQAQRL-----CLVEAESPPETWGLSNVDE 112  
 QY 250 FQTLTPESSCLREDHCHPQLSAELIPASWQGPCPLSTEGLPETHHLRMKRLFEFLOQANK 309  
 DB 113 YNA-KGQDRLREKEMERTAGKATLQAPLQAGAD-----KRLGEVAREE 156  
 QY 310 G---QDLPTPDQNGYHSLSEHSLLR-----MDPKHCRDNPOTQFVPAAGDIPGNTQ 358  
 DB 157 GVAEPAYPTSQSGGPAENEDEGETVKTYQASAASTAPGKYPVFPFLGEA-----EH 210  
 QY 359 ESTEEKIELTTVPVLALEESSEGCPS--EIPMEKEPGEGRISVVDVSYLEGDLPI 416  
 DB 211 QATEEK---GTE-NKADPSNPSGSHSRAWEYYSREKPKQ-----EGEAKVE 254  
 QY 417 ARPA-----CSNKLID-----YILGASDLETSSDPE----- 444  
 DB 255 AHRAGQGHPCRNABEEGGPPTFVCTGNALFKAWYRPGEDTEEBNSDSDAEEDTAQ 314  
 QY 445 -----GEDWDEAEEDGDFDSDSLSDSLDQDPEGLHWNPSFCSDVP 486  
 DB 315 TGATPHTSAFLKAWYRPGEDTEEE-----DSDSDAEEDTAQTGATPHT-SAFKAWV 367  
 QY 487 YNPQNT-----ATIQTAARI-----VPEEPDSEKOLSGKSDL-----E 521

```
Db 368 YRPGEDTEENSDDLSDAEEDTAQTGATPHTSAFLKAWVYRPGEDTEEE---NSDLSDAE 424
Qy 522 NSSQSGSLPET-----PEHSSGEEDDWS-----SADAEESKLWNSFC 560
Db 425 DTAGTATGATPHTSPFLKAWVYRPGEDTEEDTEEESENAPGDSATDSSQCLQPQRC 484
Qy 561 -----NSDDPYNPLNFKAPFOTSGENEGKCRDSTPSESIVAISECHTLSSCKVQLL 612
Db 485 LPGEKTKGRGEEP--PL-FQVAFYLPGEK-----PESPWAAPKLPLRLQRLRLF 531
Qy 613 GS-----QESCEPDSVQRDVLSSGRHVVHKKVTFLEEVTEYYIS-----GDEDRKGWPE 663
Db 532 KAPTRDQDPEIP-----LKARKVHFAEKVTVHFLAVWAGPAQAARRGPWE 576
Qy 664 EFARDGCRFOKRIQIETEDAIGYCLTFEHRERMFNLQ 700
Db 577 QFARDRSRFPARRIAQAEKLGPLYLTPDSRARAWRLR 613

RESULT 4
US-09-052-753B-3
; Sequence 3, Application US/09052753B
; Patent No. 6472520
; GENERAL INFORMATION:
; APPLICANT: Paul B. Fisher
; TITLE OF INVENTION: Progression Elevated Gene-3 and Uses
; FILE REFERENCE: A34608-B
; CURRENT APPLICATION NUMBER: US/09/052,753B
; CURRENT FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: PCT/US98/05793
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 08/812,818
; PRIOR FILING DATE: 1997-03-21
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 657
; TYPE: PRT
; ORGANISM: rat
US-09-052-753B-3

Query Match 6.2%; Score 236; DB 4; Length 657;
Best Local Similarity 21.8%; Pred. No. 2.3e-11;
Matches 165; Conservative 77; Mismatches 225; Indels 290; Gaps 33;

Qy 70 LLAPLPGLLQKVLINSOLFQGMFPTRWLDAGVYSALRAKLGREKPAAPTAQKSLSLQL 129
Db 21 LLSPLMGLLSRA--WSRLRGPEVPEAWL----- 46
Qy 130 DSSDSPSVTSLDWLEEGIHWOYSPDDLKLELKAGSALDPAQAFLLEQQLWGVLLPSS 189
Db 47 -----AKTVTGADQIE-----AALLTTPVSG-NLLP-- 73
Qy 190 LQSRLYSNRELGGSPGLNLTQRINDFNSVSYLLNPSYLDGPPRLEVSYQNSDGNSEYVG 249
Db 74 -----HGETESGSPQSQAAQL-----CLVEAESPPETWGLSNVDE 112
Qy 250 FQTLTPSESLREDHCHPOPIASAILPASWQCPLPSTEGLEPIHILRMKLEFLQOANK 309
Db 113 YNA-KPGQDRLREKEMERTAGKATLQAPAGQAD-----KRLGEVVAREE 156
Qy 310 G---QDLPTPDQDNGYHSLSEHSLSL-----MDPKHCRDNPTQFVPAAGDIPGNTQ 358
Db 157 GVAEPAYTISQLEGGPAENEDEGETVKYQASASIAIGYKPSFVPFLGEA-----EH 210
Qy 359 ESTERKIBLLTTEVPLALEEBSGPCPSS--EIPMEKEPGEGRISVVYDYSYLEGDLPI 416
Db 211 QATEEK---GTE-NKADPSPSPSSGSHRAWYYSREKPKQ-----EGEAKVE 254
Qy 417 ARPA-----CSNKLLD-----YILGGASSDLETSSDPE-----444
Db 255 AHRAGQGHPCRNAEAEEGGPETTFVCTGNAFILKAWVYRPGEDTEEDNSDSDAEEDTAQ 314
```

```
Qy 445 -----GEDWDEAEDDGFDSLSLSDSDLEQDPREGLHLWNSFCSDVP 486
Db 315 TGATPHTSAFLKAWVYRPGEDTEEE-----DSDSDSAEEDTAQTGATPHT-SAFKAWV 367
Qy 487 YNPQNFT-----ATIQTAARI---VPEEPSSEKDLGSKSOL-----E 521
Db 368 YRPGEDTEENSDDLSDAEEDTAQTGATPHTSAFLKAWVYRPGEDTEEE---NSDLSDAE 424
Qy 522 NSSQSGSLPET-----PEHSSGEEDDWS-----SADAEESKLWNSFC 560
Db 425 DTAGTATGATPHTSPFLKAWVYRPGEDTEEDTEEESENAPGDSATDSSQCLQPQRC 484
Qy 561 -----NSDDPYNPLNFKAPFOTSGENEGKCRDSTPSESIVAISECHTLSSCKVQLL 612
Db 485 LPGEKTKGRGEEP--PL-FQVAFYLPGEK-----PESPWAAPKLPLRLQRLRLF 531
Qy 613 GS-----QESCEPDSVQRDVLSSGRHVVHKKVTFLEEVTEYYIS-----GDEDRKGWPE 663
Db 532 KAPTRDQDPEIP-----LKARKVHFAEKVTVHFLAVWAGPAQAARRGPWE 576
Qy 664 EFARDGCRFOKRIQIETEDAIGYCLTFEHRERMFNLQ 700
Db 577 QFARDRSRFPARRIAQAEKLGPLYLTPDSRARAWRLR 613

RESULT 5
US-09-052-753B-7
; Sequence 7, Application US/09052753B
; Patent No. 6472520
; GENERAL INFORMATION:
; APPLICANT: Paul B. Fisher
; TITLE OF INVENTION: Progression Elevated Gene-3 and Uses
; FILE REFERENCE: A34608-B
; CURRENT APPLICATION NUMBER: US/09/052,753B
; CURRENT FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: PCT/US98/05793
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 08/812,818
; PRIOR FILING DATE: 1997-03-21
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 578
; TYPE: PRT
; ORGANISM: human
US-09-052-753B-7

Query Match 5.8%; Score 219; DB 4; Length 578;
Best Local Similarity 21.7%; Pred. No. 5.5e-10;
Matches 155; Conservative 79; Mismatches 223; Indels 258; Gaps 34;

Qy 53 SAQPETVSVTWTKLLS-QLLAPLPGLLQKVLINSOLFQGMFPTRWL--DFAGV----- 102
Db 4 SPRPE-HVLHKEKAHSEFYLLSPLMGFLSRA--WSRLRGPEVSEAWLAETVAGNFIQADA 60
Qy 103 -----YSALRAKLGREKPA-APTAOKSLSSQLDSSDSPSVTSLDWLEEGIHWOY 151
Db 61 LLTPPPVSENHLPURETEGNGTPEWSKAAQRLCLDVEAQSSPPKT-----WGL 108
Qy 152 SPDDLKLELKAGSALDPAQAFLLEQQLWGVLLPSSLQSRLYSNRELGGSPGLNLIQ 211
Db 109 SDID---EHNGK-----PGQDGLREQEVEHTAGLP-TLQ-----PLHLQ 143
Qy 212 RIDNFSVSYLLNPSYLDGPPRLEVSYQNSDGNSEVVGFTLTPSESLREDHCHPOPLS 271
Db 144 GADK-----KVGEVVAAREGVSALA----- 163
Qy 272 AELIPAS-WQCCP---PLSTEGLEPIHILRMKLEFLQOANKGODLPTPDQDNGYHSL 327
Db 164 ---YPTSHWEGGPAEDIEDTETVKKAH-----QASASIAIP-----GY----- 198
```



```
; ATTORNEY/AGENT INFORMATION:
; NAME: Chan, Albert Wai-Kit
; REGISTRATION NUMBER: 36,479
; REFERENCE/DOCKET NUMBER: 51523
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 590 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-821-818-2

Query Match      5.3%; Score 201; DB 3; Length 590;
Best Local Similarity 23.9%; Pred. No. 2e-08;
Matches 142; Conservative 56; Mismatches 210; Indels 186; Gaps 29;

QY 245 SEVVGFTLTPESSCLREDHCHPOPLSAEL-----IPASWQCPCPLSTEGLEPIHILRM 298
Db 23 SPLMGF--LSRAWSRLRVEA-PEPWAETVTGADQIEADAHAPPL-----VPE-NHPPQ 74
QY 299 KRLEFLOQANKGQDLTPDQDNGYHSLSEEHSLRLMDPKHCRDNP-----TOFVP--- 348
Db 75 GEAEESGTPPEGKAAQGCPLDVQANSSPPETIGLSDDDKQGDGPREQGRAHTAGLPILL 134
QY 349 -----AAGDIPGNTQESTEEKIELLTTEVPLALEEESPSEGCPSSE----- 389
Db 135 SPGLQSAKSLGEVVAEGEVT-----LAY-PTSHWEGCPSSEEDGETVKKA 182
QY 390 -----IPMEKE-----PGEGRISVDYSVLEGLDPLISARPACSNKLDIVLGASSDL 437
Db 183 FRASADSPGHKSSSTVYCPGEAEHQATEKQENKADPPSPS-----GSHRAW 232
QY 438 ETSDPGEQDWE-----EAEDDGFDSLSLSDLEQDPEGLHLWNSFCV 484
Db 233 EYCSKQEGEADPEPHRAGKYQLCQNAEAEAEAEKVSLSVSS-----GNAFLKA 282
QY 485 DPYNPQNT-----ATIQTAARIVPEEP-----SDSEKDLGSKDLE 521
Db 283 WYVRPGEDTDDSDWGSAAEEGKALSPTSPEHDFLKAWVYRPGEDTDD--DDSDWG 340
QY 522 NSSQSG---SLPETPEHS-----SGE--EDDWESSADEAES-----L 553
Db 341 SABEEGKALSPTSPEHDFLKAWVYRPGEDTDDQDSDWGSAAEKDGLAQTFATHTSAFL 400
QY 554 KLWNSFCNSDDPN-----PLNFKA--PFQTSGENEKGCRDSTPSESIVASEC-HT 603
Db 401 KTW-VCCPGEDTDDDCVVVVPEDSEAADPKSPSHEAQC-----LPGEQTEGLVEAEHS 455
QY 604 LLSCKVQLLGSQES-----ECPDSVQRDVL-----SGRHTHVKKKVTFLBEVTEY 650
Db 456 LFQVAFYLPGEKPAPWTPAKPLPLRLQRLTLRLTPTQDQDPTPLRARKVHFSENVTVH 515
QY 651 YIS-----GDEDRKGPWEFEARGCRFQKRIQETEDAIGYCLTFEHERMFNRL 699
Db 516 FLAVWAGPAQAARRGFWEQQLARDRSFARRIAQAEKGLPYLTPAFRARAARL 569

RESULT 8
US-09-052-753B-2
; Sequence 2, Application US/09052753B
; Patent No. 6472520
; GENERAL INFORMATION:
; APPLICANT: Paul B. Fisher
; TITLE OF INVENTION: Progression Elevated Gene-3 and Uses
; FILE REFERENCE: A34608-B
; CURRENT APPLICATION NUMBER: US/09/052,753B
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: PCT/US98/05793
```

```
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 08/812,818
; PRIOR FILING DATE: 1997-03-21
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 590
; TYPE: PRT
; ORGANISM: rat
US-09-052-753B-2

Query Match      5.3%; Score 201; DB 4; Length 590;
Best Local Similarity 23.9%; Pred. No. 2e-08;
Matches 142; Conservative 56; Mismatches 210; Indels 186; Gaps 29;

QY 245 SEVVGFTLTPESSCLREDHCHPOPLSAEL-----IPASWQCPCPLSTEGLEPIHILRM 298
Db 23 SPLMGF--LSRAWSRLRVEA-PEPWAETVTGADQIEADAHAPPL-----VPE-NHPPQ 74
QY 299 KRLEFLOQANKGQDLTPDQDNGYHSLSEEHSLRLMDPKHCRDNP-----TOFVP--- 348
Db 75 GEAEESGTPPEGKAAQGCPLDVQANSSPPETIGLSDDDKQGDGPREQGRAHTAGLPILL 134
QY 349 -----AAGDIPGNTQESTEEKIELLTTEVPLALEEESPSEGCPSSE----- 389
Db 135 SPGLQSAKSLGEVVAEGEVT-----LAY-PTSHWEGCPSSEEDGETVKKA 182
QY 390 -----IPMEKE-----PGEGRISVDYSVLEGLDPLISARPACSNKLDIVLGASSDL 437
Db 183 FRASADSPGHKSSSTVYCPGEAEHQATEKQENKADPPSPS-----GSHRAW 232
QY 438 ETSDPGEQDWE-----EAEDDGFDSLSLSDLEQDPEGLHLWNSFCV 484
Db 233 EYCSKQEGEADPEPHRAGKYQLCQNAEAEAEAEKVSLSVSS-----GNAFLKA 282
QY 485 DPYNPQNT-----ATIQTAARIVPEEP-----SDSEKDLGSKDLE 521
Db 283 WYVRPGEDTDDSDWGSAAEEGKALSPTSPEHDFLKAWVYRPGEDTDD--DDSDWG 340
QY 522 NSSQSG---SLPETPEHS-----SGE--EDDWESSADEAES-----L 553
Db 341 SABEEGKALSPTSPEHDFLKAWVYRPGEDTDDQDSDWGSAAEKDGLAQTFATHTSAFL 400
QY 554 KLWNSFCNSDDPN-----PLNFKA--PFQTSGENEKGCRDSTPSESIVASEC-HT 603
Db 401 KTW-VCCPGEDTDDDCVVVVPEDSEAADPKSPSHEAQC-----LPGEQTEGLVEAEHS 455
QY 604 LLSCKVQLLGSQES-----ECPDSVQRDVL-----SGRHTHVKKKVTFLBEVTEY 650
Db 456 LFQVAFYLPGEKPAPWTPAKPLPLRLQRLTLRLTPTQDQDPTPLRARKVHFSENVTVH 515
QY 651 YIS-----GDEDRKGPWEFEARGCRFQKRIQETEDAIGYCLTFEHERMFNRL 699
Db 516 FLAVWAGPAQAARRGFWEQQLARDRSFARRIAQAEKGLPYLTPAFRARAARL 569

RESULT 9
US-08-978-277A-4
; Sequence 4, Application US/08978277A
; Patent No. 6582956
; GENERAL INFORMATION:
; APPLICANT: Gelman, Irwin H.
; TITLE OF INVENTION: TUMOR SUPPRESSOR GENE
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10112-0228
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
```

COMPUTER: IBM Compatible	4.4%; Score 167; DB 4; Length 1596;	
OPERATING SYSTEM: DOS	Best Local Similarity 19.8%; Pred. No. 8.4e-05;	
SOFTWARE: FastSeq Version 1.5	Matches 151; Conservative 112; Mismatches 286; Indels 214; Gaps 35;	
CURRENT APPLICATION DATA:		
APPLICATION NUMBER: US/08/978,277A		
FILING DATE:		
CLASSIFICATION: 514		
PRIOR APPLICATION DATA:		
APPLICATION NUMBER: 08/665,401		
FILING DATE: 18-JUN-1996		
ATTORNEY/AGENT INFORMATION:		
NAME: Clark, Richard S		
REGISTRATION NUMBER: 26,154		
REFERENCE/DOCKET NUMBER: A30558 - 165/34008		
TELEPHONE: 212-408-2558		
TELEFAX: 212-765-2519		
TELEX:		
INFORMATION FOR SEQ ID NO: 4:		
SEQUENCE CHARACTERISTICS:		
LENGTH: 1596 amino acids		
TYPE: amino acid		
STRANDEDNESS: single		
TOPOLOGY: linear		
MOLECULE TYPE: peptide		
HYPOTHETICAL: NO		
ANTI-SENSE: NO		
FRAGMENT TYPE: internal		
ORIGINAL SOURCE:		
US-08-978-277A-4		

COMPUTER: IBM Compatible	4.1%; Score 156.5; DB 3; Length 1503;	
OPERATING SYSTEM: DOS	Best Local Similarity 20.4%; Pred. No. 0.00061;	
SOFTWARE: FastSeq Version 1.5	Matches 147; Conservative 82; Mismatches 249; Indels 243; Gaps 33;	
CURRENT APPLICATION DATA:		
APPLICATION NUMBER: US/08/978,277A		
FILING DATE:		
CLASSIFICATION: 514		
PRIOR APPLICATION DATA:		
APPLICATION NUMBER: 08/665,401		
FILING DATE: 18-JUN-1996		
ATTORNEY/AGENT INFORMATION:		
NAME: Clark, Richard S		
REGISTRATION NUMBER: 26,154		
REFERENCE/DOCKET NUMBER: A30558 - 165/34008		
TELEPHONE: 212-408-2558		
TELEFAX: 212-765-2519		
TELEX:		
INFORMATION FOR SEQ ID NO: 4:		
SEQUENCE CHARACTERISTICS:		
LENGTH: 1503 amino acids		
TYPE: amino acid		
STRANDEDNESS: single		
TOPOLOGY: linear		
MOLECULE TYPE: Protein		
US-08-976-255-14		

COMPUTER: IBM Compatible	4.1%; Score 156.5; DB 3; Length 1503;	
OPERATING SYSTEM: DOS	Best Local Similarity 20.4%; Pred. No. 0.00061;	
SOFTWARE: FastSeq Version 1.5	Matches 147; Conservative 82; Mismatches 249; Indels 243; Gaps 33;	
CURRENT APPLICATION DATA:		
APPLICATION NUMBER: US/08/978,277A		
FILING DATE:		
CLASSIFICATION: 514		
PRIOR APPLICATION DATA:		
APPLICATION NUMBER: 08/665,401		
FILING DATE: 18-JUN-1996		
ATTORNEY/AGENT INFORMATION:		
NAME: Clark, Richard S		
REGISTRATION NUMBER: 26,154		
REFERENCE/DOCKET NUMBER: A30558 - 165/34008		
TELEPHONE: 212-408-2558		
TELEFAX: 212-765-2519		
TELEX:		
INFORMATION FOR SEQ ID NO: 4:		
SEQUENCE CHARACTERISTICS:		
LENGTH: 1503 amino acids		
TYPE: amino acid		
STRANDEDNESS: single		
TOPOLOGY: linear		
MOLECULE TYPE: Protein		
US-08-976-255-14		

COMPUTER: IBM Compatible	4.1%; Score 156.5; DB 3; Length 1503;	
OPERATING SYSTEM: DOS	Best Local Similarity 20.4%; Pred. No. 0.00061;	
SOFTWARE: FastSeq Version 1.5	Matches 147; Conservative 82; Mismatches 249; Indels 243; Gaps 33;	
CURRENT APPLICATION DATA:		
APPLICATION NUMBER: US/08/978,277A		
FILING DATE:		
CLASSIFICATION: 514		
PRIOR APPLICATION DATA:		
APPLICATION NUMBER: 08/665,401		
FILING DATE: 18-JUN-1996		
ATTORNEY/AGENT INFORMATION:		
NAME: Clark, Richard S		
REGISTRATION NUMBER: 26,154		
REFERENCE/DOCKET NUMBER: A30558 - 165/34008		
TELEPHONE: 212-408-2558		
TELEFAX: 212-765-2519		
TELEX:		
INFORMATION FOR SEQ ID NO: 4:		
SEQUENCE CHARACTERISTICS:		
LENGTH: 1503 amino acids		
TYPE: amino acid		
STRANDEDNESS: single		
TOPOLOGY: linear		
MOLECULE TYPE: Protein		
US-08-976-255-14		

COMPUTER: IBM Compatible	4.1%; Score 156.5; DB 3; Length 1503;	
OPERATING SYSTEM: DOS	Best Local Similarity 20.4%; Pred. No. 0.00061;	
SOFTWARE: FastSeq Version 1.5	Matches 147; Conservative 82; Mismatches 249; Indels 243; Gaps 33;	
CURRENT APPLICATION DATA:		
APPLICATION NUMBER: US/08/978,277A		
FILING DATE:		
CLASSIFICATION: 514		
PRIOR APPLICATION DATA:		
APPLICATION NUMBER: 08/665,401		
FILING DATE: 18-JUN-1996		
ATTORNEY/AGENT INFORMATION:		
NAME: Clark, Richard S		
REGISTRATION NUMBER: 26,154		
REFERENCE/DOCKET NUMBER: A30558 - 165/34008		
TELEPHONE: 212-408-2558		
TELEFAX: 212-765-2519		
TELEX:		
INFORMATION FOR SEQ ID NO: 4:		
SEQUENCE CHARACTERISTICS:		
LENGTH: 1503 amino acids		
TYPE: amino acid		
STRANDEDNESS: single		
TOPOLOGY: linear		
MOLECULE TYPE: Protein		
US-08-976-255-14		

COMPUTER: IBM Compatible	4.1%; Score 156.5; DB 3; Length 1503;	
OPERATING SYSTEM: DOS	Best Local Similarity 20.4%; Pred. No. 0.0	

```
QY 146 GIHWQYSPDLKELKAKGSSALDPAQAFLLEQQLWGVLLPSSLSRLYSNRELGSPPS 205
Db 911 RVSGSSLPELQGEHLNKPFSDEHHSHR-RLKKNLEAVETLQ-----LNSKDAAKEA 962
QY 206 GPLNIQRDNFNSVWYLLNPSYLDLCPRLVSYQNSDGNSEV-----247
Db 963 GUVSALSDDSTSDQS-LLEDLSLAFPPASEPSLETDPDSLESVDVHALLDSLGSHTPQKL 1021
QY 248 -----VGQTLTPESSCLREDHCHPOPL-SABLIPAS-----WQCAPP-----LSTE 288
Db 1022 VPPDKPADSGYETENLESP---EWLHPAPEGTADEPATTDGCGHSGLPPNPVIVISDA 1078
QY 289 GLPEIHLHLMKLEFLQANKGODLTPD-----QDNGVHLSLEEHSLLRMDPKH 338
Db 1079 G-----DGHRTGV-TPTFTAGSQGYRDSAYFS-----1107
QY 339 CRDNPQVFPAAAGDIPGNTQSTEBEKIELLTTEVPL---ALEEESPS---EGCPSSEIPME 393
Db 1108 ---DNDSEPKRSEVPG-----TSPSALVLVQEQPLPEPVLPEQSPAAQDSCLEAR---K 1157
QY 394 KPEGGRISVDVSYLGEDLPISARPACSNKLIIDYILGASDLETSDPDGEDW-----448
Db 1158 SQPDESCLSALHNS---SDLELRATPEPAQT-----GVPQQVHPTDEASSPWSVLNA 1207
QY 449 -----DREAEDD-----GFDSDSL-----DSDLQDQPEGLHLMNSFCSDVPYN 488
Db 1208 ELSSGDDTETQDRRECTLASTGTNELLATYNSALDKSLSHSEG-----1253
QY 489 PONTATTIQTAAARIVPEBPSDEKDLGSKSDLENSQSGSLPTEPHSGBEDWESSAD 548
Db 1254 -----PKLKEPDIEGKY-LGKLGVSGLDLSDGMDADEED-ENSDD 1293
QY 549 EASLKLWNSFCNSDDPNPLNFKAPPQTSSENEKGCGRDSTKTPSESIVASECHTLLSCK 608
Db 1294 SDEDLRAFNLSLSESEDETEHPVPIILSNEDGRHLR-----1331
QY 609 VQLGQSSECPDSVORDVLGGRHTRVKKKATVLEBTVYISGDEDRKGPWPEFARD 668
Db 1332 -SLLKPTAANAPDLPEB-----WKKEKAVTFDDVTYILP-----DQETPKELGPC 1379
QY 669 G 669
Db 1380 G 1380

RESULT 11
US-08-993-118-7
; Sequence 7, Application US/08993118
; Patent No. 5997872
; GENERAL INFORMATION:
; APPLICANT: LUCAS, Sophie;
; APPLICANT: DE SMET, Charles;
; APPLICANT: BOON-FALLEUR, Thierry
; TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULE CODING FOR TUMOR
; TITLE OF INVENTION: REJECTION ANTIGEN PRECURSOR MAGE-C1 AND USES
; TITLE OF INVENTION: THEREOF
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 360 kb storage
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/993,118
; FILING DATE:
```

## CLASSIFICATION:

```
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/845,528
; FILING DATE: April 25, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Mary Anne Schofield
; REGISTRATION NUMBER: 36,669
; REFERENCE/DOCKET NUMBER: LUD 5455
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1142
; TYPE: amino acids
; STRANDEDNESS: single-stranded
; TOPOLOGY: linear
; US-08-993-118-7
```

Query Match 3.9%; Score 148; DB 2; Length 1142;

Best Local Similarity 20.7%; Pred. No. 0.0021;

Matches 151; Conservative 90; Mismatches 257; Indels 230; Gaps 41;

```
QY 26 PRSQA---GSSKFPPTPLGPNSENGNPTLLS--SAQPETRVSYMTKLLSQLL-APLPGLLQ 79
Db 288 PERTQSTFEFGAQSPLQIPVSSSSSSTLLSILFQSSPERTQSTFEGFPQSLQIPIMTSSFS 347
QY 80 KVLWSOLFQGMPTRWLDFAGVYSALRKAGREKPAAPTAQKSLSSQLQDSSDPSVTSP 139
Db 348 STLL--SIFQSSPESAQSTFEGF-----PQSP-----LQIPGS-PSFSST 384
QY 140 LDMLEEGHWOYSP-----PDLKLELKAKGSALDPAQAFLLEQQLWGVLLPSS 189
Db 385 LLSL-----FQSSPERTHSTFEGFPQSPLOI-----PMTSSF-----SST 419
QY 190 LQSLYNSNRELGSS-----PSGFLNIQRIDNFSVSYLL-----NPSYLDCCPR--LE 235
Db 420 LLSTLQSSPESAQSAPEGFPQSPLOIPVSSSFSYTLSSLFQSSPERTQSTFEGFPQSPLO 479
QY 236 VSVQNSDGNSEVVGFTLTPE-----SSCLREDHCHPO-----PLSALIPA--277
Db 480 IPVSSSSSSTLLSLFQSSPECTQSTFEGFPQSPLOIPQSPPEGENTHSPPL--QIVPSLP 537
QY 278 SWQG-----CPPLSTEGLEPEIHLHLMKLEFLQANKGODLPT-----PDQDNGYHS 324
Db 538 EWEDLSPHYFPQSPQGEDSLSP-HY-----FPQSPQGEDSLSPHYFPQSPQGEDS 589
QY 325 LEEHSLLRMDPKHCRD--NPTQF--VPAAGDIPGNTQSTEEKIELLTTEVPALAE--378
Db 590 LSPHY--FPQSPQGEDSMSPLYFPQSPLOGE--EFQSSILQSPVSISSSTSPSSLPQSF 644
QY 379 ----ESPSEGCPSSEIPEMEKEPGEGRISVDVSYLGEDLPISARPACSNKLIIDYILGAS 434
Db 645 PSSQSPPEGVQSPHSPQSPPEG-----MHQSQSPLOS-----678
QY 435 SDLETSDPDGEDWDEAEDDGFDSDSLSDSLEQDP-EG-----LHL-----WNSF 481
Db 679 ----PESAPGED-----SLSPLOIPQSPLEGEDSLSLHFPQSPPEWE--718
QY 482 CSVDYPNPQNTATTIQTAAARIVPEEPSDEKDLGSKSDLENSQSGSLPTEPHSGBED 541
Db 719 ---DSLSPHFPQ-----FPPQGEDFQSSLPVSISSSSTSLSPQSPFSPSPQSP 767
QY 542 DWESSADEAESLKLWNSF-----CNSDDPYNPLN--FKAPQTSSENEKGCGRDSTPS 592
Db 768 EGPAQSPLQRPVSSFFSYTLASLLQSHSPPEGFPQSPLOSP-----VSSFPS 819
QY 593 ESIVASECHTLLSCKVQLLGQSQSECPDS-VORDVLGGRHTRVKKKATVLEFEE---V 647
Db 820 STSSSLQSSPVSFPSTSSSLSKSPESPLQSPVISFSSSTLSL-----PFSESSSPV 875
QY 648 TYYIISGD 655
|||
```

Db 876 DEYTSDD 883

RESULT 12

US-08-845-528C-7

Sequence 7, Application US/08845528C

Patent No. 6027924

GENERAL INFORMATION:

APPLICANT: LUCAS, Sophie;

APPLICANT: DE SMET, Charles;

APPLICANT: BOON-FALLEUR, Thierry

TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULE CODING FOR TUMOR

TITLE OF INVENTION: REJECTION ANTIGEN PRECURSOR MAGE-C1 AND USES

TITLE OF INVENTION: THEREOF

NUMBER OF SEQUENCES: 14

CORRESPONDENCE ADDRESS:

ADDRESSEE: Felfe & Lynch

STREET: 805 Third Avenue

CITY: New York City

STATE: New York

COUNTRY: USA

ZIP: 10022

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.5 inch, 360 kb storage

COMPUTER: IBM PS/2

OPERATING SYSTEM: PC-DOS

SOFTWARE: Wordperfect

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/845,528C

FILING DATE: April 25, 1997

CLASSIFICATION: 4335

ATTORNEY/AGENT INFORMATION:

NAME: Mary Anne Schofield

REGISTRATION NUMBER: 36,669

REFERENCE/DOCKET NUMBER: LUD 5455

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 688-9200

TELEFAX: (212) 838-3884

INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:

LENGTH: 1142

TYPE: amino acids

STRANDEDNESS: single-stranded

TOPOLOGY: linear

US-08-845-528C-7

Query Match 3.9%; Score 148; DB 3; Length 1142;

Best Local Similarity 20.7%; Pred. No. 0.0021;

Matches 151; Conservative 90; Mismatches 257; Indels 230; Gaps 41;

Qy 26 PRRSQA---GSSKPTPLGPENSGNPTLLS---SAQPTRVSYWTKLLSLLQ-L-APLPGLLQ 79

Db 288 PERTQSTFEFGAQPQLQIPVSSSSSSLLSLFQSSPERTQSTFEGFPQSLQIPMTSSFS 347

Qy 80 KVLWSQLFGMFTRWLDGAVTYSALRAKGRKPAAPTAQKSLSLQDSSDPSTSP 139

Db 348 STLL--SIFQSSPSAQSTFEGF-----PQSP-----LQIPGS--PSFSST 384

Qy 140 LDMLEEGHMQYSP-----PDLKLELKAGKSALDPAQAFLLEQQLMGVLLPSS 189

Db 385 LLSL-----FOSSPERTHTSTFEGFPQSPQLQI-----PMTSSF-----SST 419

Qy 190 LQSRLYSNRELKSS-----PSGPNLNIQIDNFSVVSYLL-----NPSYLDCCPR--LE 235

Db 420 LLSLQSSPSAQAFEGFPQSPQLQIPVSSSSSYLLSLFQSSPERTQSTFEGFPQSPQLQ 479

Qy 236 VSYQNSGNSVVGQFQTLTP-----SSCLREDHCHIQ-----PLSALIPA-- 277

Db 480 IPVSSSSSSSTLLSLFQSSPERTQSTFEGFPQSPQLQIPQSPPEGENTHSLP--QIVPSLP 537

Qy 278 SWQG-----CPPLSTGCLPIHHLKRLFLQKANGQDLPT-----PDQDNGYHS 324

Db 538 EWEDSLSPHYFPQSPQGEDSLSP-HY-----FPQSPQGEDSLSPHYFPQSPQGEDS 589

Qy 325 LEEHSLLRMDPKHCRD--NPTQF--VPAAGDIPGNTQSTETKEIKELLTTTEVFLALEE-- 378

Db 590 LSPHY--FPQSPQGEDSMSPLYFPQSPQLQGE---EFQSSLQSPVSISSSTPSSLPQSF 644

Qy 379 ----ESPSEGCPSSEIPMEKEPCEGRISVVVDYSLYEGDLPISARPAKSNKLDIYILGGAS 434

Db 645 PESSQSPPEGVPQSPHSPQSPPEG-----MHSQSPLOQ----- 678

Qy 435 SDLETSDPEGEDWDEEAEDDGFDSLSLSDSLEQDP-EG-----LHL-----WNSF 481

Db 679 ----PESAEGED-----SLSPLOIPQSPLEGEDSLSLHFPQSPPEWE-- 718

Qy 482 CSVDPPNPQNTATIOTAAARIVPEPDSSEKOLSGKSDLENSQSQSLPPTPHHSGEED 541

Db 719 ---DSLSPHLFPQ-----FPQGEDFQSSLQSPVSISSSTSLSPQSPFSPSP 767

Qy 542 DMESSADEAESLKWNSF-----CNSDDPYNPLN--FKAPFQTSGENEKGCRCDSKTPS 592

Db 768 EGPAQSPQRPQVSSFFSYTLASLLQSSHSPQSPPEGPAQSPLOQSP-----VSSFPS 819

Qy 593 ESIVAISECHTLILCKVQLLGSECEPDS-VQDVLSGGRHTRVKKRVTFLEE----V 647

Db 820 STSSSLQSPVSSFPSSSTSSLSKSPSPLOQSPVIFSSSTSLs-----PFSESSSPV 875

Qy 648 TEYIISGD 655

Db 876 DEYTSDD 883

RESULT 13

US-09-066-281B-7

Sequence 7, Application US/09066281B

Patent No. 6475783

GENERAL INFORMATION:

APPLICANT: LUCAS, Sophie; DE SMET, Charles; BOON-FALLEUR, Thierry

TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULE CODING

TITLE OF INVENTION: FOR TUMOR REJECTION ANTIGEN PRECURSOR MAGE-C1 AND MAGE-C2

TITLE OF INVENTION: AND USES THEREOF

NUMBER OF SEQUENCES: 20

CORRESPONDENCE ADDRESS:

ADDRESSEE: Fulbright & Jaworski L.L.P.

STREET: 666 Fifth Avenue

CITY: New York City

STATE: New York

COUNTRY: USA

ZIP: 10103

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.5 inch, 360 kb storage

COMPUTER: IBM PS/2

OPERATING SYSTEM: PC-DOS

SOFTWARE: Wordperfect

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/066,281B

FILING DATE: April 24, 1998

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/845,528

FILING DATE: April 25, 1997

ATTORNEY/AGENT INFORMATION:

NAME: Mary Anne Schofield

REGISTRATION NUMBER: 36,669

REFERENCE/DOCKET NUMBER: LUD 5455.2 US - JEL/MAS

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 318-3100

TELEFAX: (212) 752-5958

INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:

LENGTH: 1142

TYPE: amino acids

STRANDEDNESS: single-stranded

TOPOLOGY: linear

US-09-066-281B-7



```
Query Match      3.9%; Score 148; DB 4; Length 1142;
Best Local Similarity 20.7%; Pred. No. 0.0021;
Matches 151; Conservative 90; Mismatches 257; Indels 230; Gaps 41;

QY 26 PRRSQA---GSKFTPLGPNNGNPTLLS--SAQPTRVSYWTKLSQLL--APLGLLQ 79
DB 288 PERTQSTFEQAQSLQIPVSSSSSTLLSLFQSSPERTQSTFEQFQSLQIPMTSSFS 347
QY 80 KVLWSQLFGMPTRWLDFAGVYSALRAKREKPAAPTAKSLSSLOLDSSDPSVTSP 139
DB 348 STLL--SIFQSSPESAQSTFEF-----PQSP-----LQIPGS--PSFSST 384
QY 140 LDWLEEGIHQYSP-----PDLKLELKAKGSALDPAQAFLLEQQLGWVLLPSS 189
DB 385 LLSL-----FQSSPERTSTFEFQSPQLQI-----PMTSSF-----SST 419
QY 190 LQSRLYSNRELGS-----PSGPLNIQRIINFVSVYLL-----NPSYLDCTPR--LE 235
DB 420 LLSILQSSPESAQSAFEGFPQSPQLQIPVSSSFSYTLLSLFQSSPERTQSTFEFQSPQLQ 479
QY 236 VSYQNSDGNSEVVGFTLTPE-----SSCLREDHCHPQ-----PLSAELIPA-- 277
DB 480 IPVSSSSSTLLSLFQSSPERTQSTFEFQSPQLQIPQSPPEGENTHSP--L--QIVPSLP 537
QY 278 SWQG-----CPPLSTEGLEPEIHLRMKRLLEFLOQANKGQDLPT-----PDQDNGYHS 324
DB 538 EWEDSLSPHYFPQSPQGEDSLSP-HY-----FPQSPQGEDSLSPHYFPQSPQGEDS 589
QY 325 LEEHSLLRMDPKHCRD--NPTQF--VPAAGDIPGNTQESTEEKIELLTTEVPALAE-- 378
DB 590 LSPHY--FPQSPQGEDSMSPLYPFQSPLOQE---EFQSSLQSPVSISSSTPSSLPQSF 644
QY 379 ----ESPSEGCPSSEIPMEKEPCEGRISVVDYSYLEGDLPI SARPACSKNKLIDYLGAS 434
DB 645 PESSQSPPEGVPQSPHSPQSPPEG-----MHSQSPLOS----- 678
QY 435 SDLETSSDPGEDWDDEAEADDGFDSDSLSDSDLEQDP--EG-----LHL-----WNSF 481
DB 679 ----PESAPEGED-----SLSPQLQIPQSPLEGEDSLSLHFPQSPPEWE-- 718
QY 482 CSVDYPNQFTATIQTAAIRIVPEEPSDSKDLGKSDLENSQSGSLPPTPEHSSGEED 541
DB 679 ----PESAPEGED-----SLSPQLQIPQSPLEGEDSLSLHFPQSPPEWE-- 718
QY 482 CSVDYPNQFTATIQTAAIRIVPEEPSDSKDLGKSDLENSQSGSLPPTPEHSSGEED 541
```

```
Query Match      3.9%; Score 148; DB 4; Length 1142;
Best Local Similarity 20.7%; Pred. No. 0.0021;
Matches 151; Conservative 90; Mismatches 257; Indels 230; Gaps 41;

QY 26 PRRSQA---GSKFTPLGPNNGNPTLLS--SAQPTRVSYWTKLSQLL--APLGLLQ 79
DB 288 PERTQSTFEQAQSLQIPVSSSSSTLLSLFQSSPERTQSTFEQFQSLQIPMTSSFS 347
QY 80 KVLWSQLFGMPTRWLDFAGVYSALRAKREKPAAPTAKSLSSLOLDSSDPSVTSP 139
DB 348 STLL--SIFQSSPESAQSTFEF-----PQSP-----LQIPGS--PSFSST 384
QY 140 LDWLEEGIHQYSP-----PDLKLELKAKGSALDPAQAFLLEQQLGWVLLPSS 189
DB 385 LLSL-----FQSSPERTSTFEFQSPQLQI-----PMTSSF-----SST 419
QY 190 LQSRLYSNRELGS-----PSGPLNIQRIINFVSVYLL-----NPSYLDCTPR--LE 235
DB 420 LLSILQSSPESAQSAFEGFPQSPQLQIPVSSSFSYTLLSLFQSSPERTQSTFEFQSPQLQ 479
QY 236 VSYQNSDGNSEVVGFTLTPE-----SSCLREDHCHPQ-----PLSAELIPA-- 277
DB 480 IPVSSSSSTLLSLFQSSPERTQSTFEFQSPQLQIPQSPPEGENTHSP--L--QIVPSLP 537
QY 278 SWQG-----CPPLSTEGLEPEIHLRMKRLLEFLOQANKGQDLPT-----PDQDNGYHS 324
DB 538 EWEDSLSPHYFPQSPQGEDSLSP-HY-----FPQSPQGEDSLSPHYFPQSPQGEDS 589
QY 325 LEEHSLLRMDPKHCRD--NPTQF--VPAAGDIPGNTQESTEEKIELLTTEVPALAE-- 378
DB 590 LSPHY--FPQSPQGEDSMSPLYPFQSPLOQE---EFQSSLQSPVSISSSTPSSLPQSF 644
QY 379 ----ESPSEGCPSSEIPMEKEPCEGRISVVDYSYLEGDLPI SARPACSKNKLIDYLGAS 434
DB 645 PESSQSPPEGVPQSPHSPQSPPEG-----MHSQSPLOS----- 678
QY 435 SDLETSSDPGEDWDDEAEADDGFDSDSLSDSDLEQDP--EG-----LHL-----WNSF 481
DB 679 ----PESAPEGED-----SLSPQLQIPQSPLEGEDSLSLHFPQSPPEWE-- 718
QY 482 CSVDYPNQFTATIQTAAIRIVPEEPSDSKDLGKSDLENSQSGSLPPTPEHSSGEED 541
DB 719 ---DSLSPHFPQ-----FPQGEDFQSSLQSPVSISSSTSLSPQSPFSPSP 767
QY 542 DWESADAEASLKLWNSF-----CNSDDPVNPLN--FKAPFOTSGENKGRCDKTPS 592
DB 768 EGPAQSPLOQPVSSPFTSLASLQSSHESQSPPEGPAQSPLOQSP-----VSSPFS 819
QY 593 ESIVAISECHTLKCKVQLLQSGSECPDS--VQRDVLGSGRHTVYKRVKVTFLBE--V 647
DB 820 STSSSLQSSPVSPPSSSTSSSLKSPSPESPLQSPVIFSSTSLSL-----PFSESSSPV 875
QY 648 TEYIISGD 655
DB 876 DEYTSDD 883
```

```
RESULT 14
US-09-468-433C-7
; Sequence 7, Application US/09468433C
; Patent No. 6680191
; GENERAL INFORMATION:
; APPLICANT: LUCAS, Sophie; BOON-PALLEUR, Thierry
; TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULES CODING FOR
; TITLE OF INVENTION: TUMOR REJECTION ANTIGEN PRECURSORS OF MEMBERS OF THE MAGE-C AN
; TITLE OF INVENTION: MAGE-B FAMILIES AND USES THEREOF
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fulbright & Jaworski L.L.P.
; STREET: 801 Pennsylvania Avenue, NW
; CITY: Washington
; STATE: District of Columbia
```

Db 719 ---DSLSPFPQ-----FPQGEDFQSSLPVSISSSTSLSPQSPSP 767  
Qy 542 DWSSADEAELKWNKF-----CNSDDPYPLN--FKAPFOTSGENEGCRDSTPS 592  
Db 768 EGPAQSLQRPVSVFFYTLASLQSHSPSPPEGPAQLQSP-----VSSFPS 819  
Qy 593 ESIVAISECHTLCKVQLLGSQSECPDS-VQRDVLGGRRHVRKRVKVFLEE---V 647  
Db 820 STSSLSQSPVSPSPSSSLKSPSPSPVLSFSSTSL-----PFSSESSPV 875  
Qy 648 TEYIISGD 655  
Db 876 DEXTSSD 883

RESULT 15  
US-09-060-410-4  
; Sequence 4, Application US/09060410  
; Patent No. 6165461  
; GENERAL INFORMATION:  
; APPLICANT: Cobb, Melanie  
; APPLICANT: Hutchinson, Michele  
; APPLICANT: Chen, Zhu  
; APPLICANT: Berman, Kevin  
; TITLE OF INVENTION: TAO PROTEIN KINASES AND METHODS OF USE  
; TITLE OF INVENTION: THEREFOR  
; NUMBER OF SEQUENCES: 26  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SEED and BERRY LLP  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/060,410  
; FILING DATE: 14-APR-1998  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Maki, David J.  
; REGISTRATION NUMBER: 31,392  
; REFERENCE/DOCKET NUMBER: 860098.421  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 622-4900  
; TELEFAX: (206) 682-6031  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 993 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-09-060-410-4

Query Match 3.7%; Score 141; DB 3; Length 993;  
Best Local Similarity 20.7%; Pred. No. 0.0068;  
Matches 106; Conservative 68; Mismatches 164; Indels 174; Gaps 24;  
Qy 283 PPL-STEGLPRIHLRMKRLFFLOQANKG-----QDLPTPDQDNGYHSL 326  
Db 223 PPLFNWAMSAHYHIAQNESPALGSHWSEYFRNFVDSCLQKIPQDRPTSE-----VLL 276  
Qy 327 BEHSLRLMDPKHCRDNPQFVPAAGIPGNQTQESTEEKIELLTTEVPLALEEESPSGCP 386  
Db 277 KHRFVLRERP-----FTVIM---DLIQRTKDAVRELDNLQYRKMKKILFQEAENG--P 324  
Qy 387 SSEIPEMEKEPGE-----GRISVVVDYSYEGDLPISARPACS---NKLIIDYILGASSDLE 438

Db 325 GABAPEEEEAEFYMHRAGTLTSLSSHVSFVMSISASSQSSSVNLSAD-----ASDNE 378  
Qy 439 TSDDPEGEDWDEAEADDGFDSSSLSDSOLEQDPPEGLHLMNSFCSV-----DP 486  
Db 379 EEEEEEEEE-EEEEEEGEPES-----REMANMQEGEHTVTSHSIIHRLPGSDNLYDDP 431  
Qy 487 YNPQ-----NFTATIQT-----ARIVPEEPSDS--E 511  
Db 432 YQEMTPGPLOPPAAPPTSTSSSARRRAYCRNRDHFATIRTASLVSRQIQEHEQDSALR 491  
Qy 512 KDLSGKSDLENSQSGLP-----ETPEHSSGEEDDWESSADEAESLKLWNSFCNSDD 564  
Db 492 EQLSGYKRMRRQHKQLLALLESRLGEREERHSRLQRELEAQ----- 533  
Qy 565 PYNPLNFKAPFOTSGE-----NEKGRCDKSTPSESIVAISECHTLCKVQLLGSQ 615  
Db 534 -----RAGFGTEAEKLAARRHQAIGKEARAAQAERKF-----QQHILGQQ 574  
Qy 616 ESECPSVQRDVLGGRRHVRKRVKVFLEEYVYISGDEDRKGPWEERFARDCGCFQKR 675  
Db 575 KKEL-----AALLEAQKTY-KLRKEQLKEEL-----QENPSTPKREKAEWLLRQKEQ 621  
Qy 676 IQ--ETEDAIGYCLTFEHRERMFNRLOGTCFK 705  
Db 622 LQOCQAEERAGLL-----RRQRQYFELQCROYK 649

Search completed: September 16, 2005, 10:47:17  
Job time : 23.202 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

# OM protein - protein search, using sw model

Run on: September 16, 2005, 10:46:08 ; Search time 243.057 Seconds  
(without alignments)  
1187.844 Million cell updates/sec

Title: US-10-650-482-2

Perfect score: 3804

Sequence: 1 MBEFTGSRKRLGRAGFRF.....RMFNRLQGTCTGKLVNLKQC 713

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1812044 seqs, 404927589 residues

Total number of hits satisfying chosen parameters: 1812044

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : Published Applications AA:\*

- 1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/1/pubpaa/US10D\_PUBCOMB.pep.\*
- 17: /cgn2\_6/ptodata/1/pubpaa/US10E\_PUBCOMB.pep.\*
- 18: /cgn2\_6/ptodata/1/pubpaa/US10F\_PUBCOMB.pep.\*
- 19: /cgn2\_6/ptodata/1/pubpaa/US11A\_PUBCOMB.pep.\*
- 20: /cgn2\_6/ptodata/1/pubpaa/US11\_NEW\_PUB.pep.\*
- 21: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*
- 22: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3804	100.0	713	16	US-10-650-482-2
2	3627	95.3	707	9	US-09-925-299-941
3	3627	95.3	707	10	US-09-925-299-941
4	2223	58.4	698	16	US-10-650-482-4
5	813	21.4	153	15	US-10-296-115-963
6	259.5	6.8	674	16	US-10-755-889-90
7	236	6.2	657	9	US-09-052-753-3
8	236	6.2	657	15	US-10-277-603-3
9	219	5.8	578	9	US-09-052-753-7
10	219	5.8	578	15	US-10-277-603-7
11	201	5.3	590	9	US-09-052-753-2

Sequence 2, Appl1	590	15	US-10-277-603-2	Sequence 2, Appl1
Sequence 4, Appl1	1596	9	US-09-902-432-4	Sequence 4, Appl1
Sequence 13, Appl1	4.2	1562	14	US-10-097-534-13
Sequence 57852, A	4.2	1562	17	US-10-450-763-57852
Sequence 1, Appl1	4.2	1585	17	US-10-487-132-1
Sequence 38864, A	4.2	1702	18	US-10-450-763-38864
Sequence 143280, A	4.1	1772	16	US-10-437-963-143280
Sequence 62, Appl1	4.1	2766	11	US-09-964-956-62
Sequence 44262, A	4.1	888	18	US-10-450-763-44262
Sequence 54269, A	4.1	888	18	US-10-450-763-54269
Sequence 18, Appl1	4.0	1098	15	US-10-161-927-18
Sequence 2, Appl1	4.0	764	14	US-10-399-140-2
Sequence 38863, A	4.0	1391	18	US-10-450-763-38863
Sequence 1773, App	3.9	787	15	US-10-108-260A-3773
Sequence 3775, App	3.9	1275	18	US-10-840-512-175
Sequence 462, App	3.9	730	14	US-10-153-668-462
Sequence 2173, App	3.9	1857	16	US-10-408-765A-2173
Sequence 448, App	3.9	428	16	US-10-451-467A-448
Sequence 4505, App	3.9	553	15	US-10-108-260A-4505
Sequence 12, Appl1	3.9	878	15	US-10-467-434-12
Sequence 247745, A	3.9	612	16	US-10-425-115-247745
Sequence 185523, A	3.9	745	16	US-10-437-963-185523
Sequence 7, Appl1	3.9	1142	13	US-10-085-108-7
Sequence 79, Appl1	3.9	1142	14	US-10-207-655-79
Sequence 7, Appl1	3.9	1142	14	US-10-160-237-7
Sequence 195416, A	3.9	1208	18	US-10-450-763-195416
Sequence 143282, A	3.9	672	16	US-10-437-963-143282
Sequence 529, App	3.9	1844	16	US-10-408-765A-529
Sequence 202963, A	3.9	3394	16	US-10-437-963-202963
Sequence 36, Appl1	3.8	672	16	US-10-437-963-36
Sequence 42484, A	3.8	835	16	US-10-755-889-36
Sequence 276711, A	3.7	460	16	US-10-767-701-42484
Sequence 5, Appl1	3.7	976	16	US-10-425-115-276711
Sequence 195, App	3.7	1572	17	US-10-487-132-5
Sequence 262, App	3.7	2132	18	US-10-840-512-195
Sequence 106033, A	3.7	406	16	US-10-451-467A-262
Sequence 4755, App	3.7	1774	16	US-10-437-963-106033
Sequence 1, Appl1	3.7	2724	17	US-10-108-260A-4755
Sequence 4, Appl1	3.7	2724	17	US-10-487-593-1
Sequence 32497, A	3.7	993	15	US-10-445-735-4
Sequence 52460, A	3.7	1714	18	US-10-450-763-32497
Sequence 7, Appl1	3.7	635	15	US-10-450-763-52460
Sequence 165014, A	3.7	1132	15	US-10-408-967-7
Sequence 363, App	3.7	3975	16	US-10-437-963-165014
Sequence 2, Appl1	3.7	1235	13	US-10-087-192-363
Sequence 36776, A	3.7	1346	9	US-09-902-432-2
Sequence 2, Appl1	3.6	1714	18	US-10-450-763-36776
Sequence 73, Appl1	3.6	1142	9	US-09-899-651-2
Sequence 54525, A	3.6	1142	16	US-10-670-472A-73
Sequence 18, Appl1	3.6	1155	18	US-10-450-763-54525
Sequence 162500, A	3.6	3067	10	US-09-949-029-18
Sequence 224762, A	3.6	572	16	US-10-437-963-162500
Sequence 2174, App	3.6	576	15	US-10-424-599-224762
Sequence 202, App	3.6	1203	16	US-10-408-765A-2174
Sequence 8157, App	3.6	1388	18	US-10-840-512-202
Sequence 43815, A	3.6	2688	17	US-10-732-923-8157
Sequence 52462, A	3.6	496	16	US-10-767-701-43815
Sequence 16, Appl1	3.6	904	18	US-10-450-763-52462
Sequence 7, Appl1	3.6	1049	14	US-10-317-835-16
Sequence 27, Appl1	3.6	1049	15	US-10-168-582-7
Sequence 27, Appl1	3.6	581	10	US-09-244-803-27
Sequence 27, Appl1	3.6	581	16	US-09-245-277-27
Sequence 11, Appl1	3.6	581	16	US-10-792-481-27
Sequence 240, App	3.6	1732	18	US-10-257-763-11
Sequence 240, App	3.6	1732	16	US-09-918-715-240
Sequence 1, Appl1	3.6	1732	16	US-10-474-794-240
Sequence 67803, A	3.6	1735	18	US-10-979-153-240
Sequence 138582, A	3.6	1735	16	US-10-333-680-1
Sequence 316809, A	3.5	747	15	US-10-425-114-67803
Sequence 217, App	3.5	811	16	US-10-437-963-138582
Sequence 105693, A	3.5	967	16	US-10-425-115-316809
	3.5	1469	18	US-10-840-512-217
	3.5	1015	16	US-10-437-963-105693

```

85 134.5 3.5 1253 14 US-10-363-798-2 Sequence 2, Appli
86 134.5 3.5 1692 17 US-10-732-923-8431 Sequence 8431, Ap
87 134.5 3.5 1969 9 US-09-839-479-71 Sequence 71, Appl
88 134.5 3.5 1969 15 US-10-376-537-72 Sequence 72, Appl
89 134.5 3.5 1969 15 US-10-376-537-72 Sequence 71, Appl
90 134.5 3.5 2468 16 US-10-755-889-615 Sequence 615, App
91 134.5 3.5 2468 17 US-10-489-740-216 Sequence 216, App
92 134.5 3.5 2519 18 US-10-450-763-46995 Sequence 46995, A
93 134.5 3.5 2622 18 US-10-450-763-36777 Sequence 36777, A
94 134.5 3.5 3529 17 US-10-732-923-8327 Sequence 8327, Ap
95 134.5 3.5 3571 17 US-10-732-923-8324 Sequence 8324, Ap
96 134.5 3.5 3574 17 US-10-732-923-8326 Sequence 8326, Ap
97 134.5 3.5 3574 17 US-10-732-923-8329 Sequence 8329, Ap
98 134 3.5 812 16 US-10-437-963-117056 Sequence 117056,
99 134 3.5 1439 16 US-10-719-993-803 Sequence 803, App
100 133.5 3.5 446 16 US-10-425-115-264340 Sequence 264340,

ALIGNMENTS

; Sequence 2, Application US/10650482
; Publication No. US20040142345A1
; GENERAL INFORMATION:
; APPLICANT: Ron, David
; APPLICANT: Jousse, Celine
; TITLE OF INVENTION: METHODS OF SCREENING TEST COMPOUNDS USING GADD34L, AN eIF2alpha-S
; FILE REFERENCE: PHOSPHATASE SUBUNIT
; CURRENT APPLICATION NUMBER: 5986/11712-US1
; CURRENT FILING DATE: 2003-08-28
; PRIOR APPLICATION NUMBER: US 60/408,679
; PRIOR FILING DATE: 2002-09-06
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 713
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-650-482-2

Query Match 100.0%; Score 3804; DB 16; Length 713;
Best Local Similarity 100.0%; Pred. No. 1.2e-278;
Matches 713; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MEPGTGSRKRLGPRAGFRFPWPPPPRRSQAGSKFPTPLGPENSGNPTLLSSAQPETRV 60
Db 1 MEPGTGSRKRLGPRAGFRFPWPPPPRRSQAGSKFPTPLGPENSGNPTLLSSAQPETRV 60

Qy 61 SYWTKLLSOLLAPLPGLLQKVLKLSQLFGMPTPLWDLFAGVYSALRALKGRKXPAPTA 120
Db 61 SYWTKLLSOLLAPLPGLLQKVLKLSQLFGMPTPLWDLFAGVYSALRALKGRKXPAPTA 120

Qy 121 QKSLSLLQDSDSPVTSPLDLESGIHQVSPDPLKLEKAKGSALDPAQAFLLEQOL 180
Db 121 QKSLSLLQDSDSPVTSPLDLESGIHQVSPDPLKLEKAKGSALDPAQAFLLEQOL 180

Qy 181 WGVLLPSSLSRLYSNRELSSPGPLNIQRIDNFSVSVLLNPSYLDLCPFRLEVSYQN 240
Db 181 WGVLLPSSLSRLYSNRELSSPGPLNIQRIDNFSVSVLLNPSYLDLCPFRLEVSYQN 240

Qy 241 SDGNSEVVGFTLTPESSCLREDHCHPOPLSAELIPASWQCCPLSTGLPEIHLRMR 300
Db 241 SDGNSEVVGFTLTPESSCLREDHCHPOPLSAELIPASWQCCPLSTGLPEIHLRMR 300

Qy 301 LEFLQANKGQDLPTPDQDNGVHSLSEEHSLRLMDPKHCRDNPTQFVPAAGDIPGNTQES 360
Db 301 LEFLQANKGQDLPTPDQDNGVHSLSEEHSLRLMDPKHCRDNPTQFVPAAGDIPGNTQES 360

Qy 361 TEEKIELLTTEVPLALEBESSEGCPSSEIPEMEKEPGEGRISVVDVSYLEGDLPIARP 420

```

```

Db 361 TEEKIELLTTEVPLALEBESSEGCPSSEIPEMEKEPGEGRISVVDVSYLEGDLPIARP 420
Qy 421 CSNKLIDYILGGASSDLETSSDDEGEDWDEAEADDGDSLSLSDSDLEODPEGLHLWNS 480
Db 421 CSNKLIDYILGGASSDLETSSDDEGEDWDEAEADDGDSLSLSDSDLEODPEGLHLWNS 480
Qy 481 FCSVDYPNPNFTATIQTAAARIIVPEEPSDEKOLSGKSDLENSSQSGSLPETPHSSGEE 540
Db 481 FCSVDYPNPNFTATIQTAAARIIVPEEPSDEKOLSGKSDLENSSQSGSLPETPHSSGEE 540
Qy 541 DDMWESSADEAESLKLWNSFCNSDDPNPLNFKAPFQTSGENEKGRDSKTPSEISVAISE 600
Db 541 DDMWESSADEAESLKLWNSFCNSDDPNPLNFKAPFQTSGENEKGRDSKTPSEISVAISE 600
Qy 601 CHTLLSCKVOLIGSQSESCPDVSQVORDVLSGGRHTHVKRKKVTFLEEVTEYISGDEDRKG 660
Db 601 CHTLLSCKVOLIGSQSESCPDVSQVORDVLSGGRHTHVKRKKVTFLEEVTEYISGDEDRKG 660
Qy 661 PWEEFARDGCRFQKRIQETEDAIGCYCLTFEHRERMFNRLOQTCFKGLNLVKQC 713
Db 661 PWEEFARDGCRFQKRIQETEDAIGCYCLTFEHRERMFNRLOQTCFKGLNLVKQC 713

```

```

RESULT 2
US-09-925-299-941
; Sequence 941, Application US/09925299
; Patent No. US20020055627A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 941
; LENGTH: 707
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: SITE
; LOCATION: (265)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (271)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (307)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; US-09-925-299-941

```

```

Query Match 95.3%; Score 3627; DB 9; Length 707;
Best Local Similarity 98.8%; Pred. No. 3e-265;
Matches 683; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

Qy 23 PFFPRRSQAGSKFPTPLGPENSGNPTLLSSAQPETRVSYWTKLLSOLLAPLPGLLQKVL 82
Db 17 PFSVRSRQAGSKFPTPLGPENSGNPTLLSSAQPETRVSYWTKLLSOLLAPLPGLLQKVL 76

Qy 83 IWQLFGMPTPLWDLFAGVYSALRALKGRKXPAPTAQKLSLQDSDSPVTSPLDW 142
Db 77 IWQLFGMPTPLWDLFAGVYSALRALKGRKXPAPTAQKLSLQDSDSPVTSPLDW 136

Qy 143 LEEGIHWQYSPDPLKLEKAKGSALDPAQAFLLEQOLWGVLLPSSLSRLYSNRELGS 202
Db 137 LEEGIHWQYSPDPLKLEKAKGSALDPAQAFLLEQOLWGVLLPSSLSRLYSNRELGS 196

Qy 203 SPGGLNIQRIDNFSVSVLLNPSYLDLCPFRLEVSYQNDCNENGVGFTLTPESSCLRE 262

```







Qy 491 NFTAQTAAARI VPEPDSSEKDLGSKDLENSQSGSLP-----ETPEH 535  
Db 395 EDT-----EEEDSDTGSAAEDEREASTTPASAFKAWVVRPGEDTEEE 443  
Qy 536 -----SSGBEDDWEASDAEASLKWNSFCNSDDPYNPLNFKAPFOTSGENEGKCHDSK 589  
Db 444 EDEDVDSKEDDSEALGAEAS-----DHP-----SHPDQSAHFRGNGYRPGK 488  
Qy 590 TPESIVA-----ISECHTLLSCKVQLLSGQ-----ESECPSVQRDVLSGGRHTH--- 635  
Db 489 ETEEEAAEDWGEAEPCEPFAVA---IYVGEKPPPPWAPPRLPLRLQRLKRPETPHDPD 546  
Qy 636 -----VKRKVTFLEEVTEYYIS-----GDEDRKGFWEFARDGCRKORRQIETEDAIYC 686  
Db 547 PETPLKARKVRFSEKVTVHFLAVWAGPAQAARQGFWEQLARDRSRFPARRIAQAQEELSPC 606  
Qy 687 LTPHRRERMFNLQCTCFKGLNVLKQ 712  
Db 607 LTPAARARAWARLURNPLAPIPALTQ 632

## RESULT 7

US-09-052-753-3  
; Sequence 3, Application US/09052753A  
; Patent No. US20010014734A1  
; GENERAL INFORMATION:  
; APPLICANT: Fisher, Paul B.  
; TITLE OF INVENTION: PROGRESSION ELLEVATED GENE-3 AND USES THEREOF  
; FILE REFERENCE: 0575/51523-B/JPW/JSG  
; CURRENT APPLICATION NUMBER: US/09/052,753A  
; CURRENT FILING DATE: 1998-03-31  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 3  
; LENGTH: 657  
; TYPE: PRT  
; ORGANISM: Rat  
US-09-052-753-3

Query Match 6.2%; Score 236; DB 9; Length 657;  
Best Local Similarity 21.8%; Pred. No. 1e-08;  
Matches 165; Conservative 77; Mismatches 225; Indels 290; Gaps 33;

Qy 70 LLAPLPGLLQKVLWISQLFGGMFPTRWLDPAQVYSALRALKGRKPAAPTAKSLSLQ 129  
Db 21 LLSPLMGLLSRA--WSRLRGPEVPEAWL----- 46  
Qy 130 DSDPSVTSPLDWLEEGIHQWYSPDLKLELKAGKSALDPAQAQAFLEQQLWGVLLPSS 189  
Db 47 ----AKTVTGADQIE-----AAALLTPTVSG-NLLP-- 73  
Qy 190 LQSLRYNRELGSPPSPLNIQRIIDNFVSVYLLNPSYLDGFPRLVSYQNSDGNSEVVG 249  
Db 74 ----HGTEESGSPESQQAQRL-----CLVEAESPPETWGLSNVDE 112  
Qy 250 FQTLTPSSCLRDHCHPQPLSAELIPASWQGCPLSTGLPEIHLHMKRLEFLQOANK 309  
Db 113 YNA-KPGQDRLREKEMERTAGKATLQAGLQAGD-----KRLGEVVAEE 156  
Qy 310 G----QDLPTPDQNGYHSLBEEHSLR-----MDPKHCRDNPOTQFVPAAGDIPGNTQ 358  
Db 157 GVAPAYPTSQLGGPAENEEDGETVTKYQASAASIAPGYKPTVPVFLGEA-----EH 210  
Qy 359 ESTEEKIELLTTEVPLALEBESSEGCPS--EIPMEKEPGEGRISVVDYSYLEGDLPI 416  
Db 211 QATEEK-----GTE-NKADPSNSPSSGSHRAWETYSREKPKQ-----EGEAKVE 254  
Qy 417 ARPA-----CSNKLID-----YILGASSDLETSSDPE----- 444  
Db 255 AHRAGQHPCNAEAESEGGPTTTCGNAPLAKAWVVRPGEDTEEDNSDSDSAEETAQ 314  
Qy 445 -----GEMWDEAEDDGDPSDLSLSDLEQDRGLHLWNSFCSDVP 486

Db 315 TGAHTPSAPLAKAWVVRPGEDTEEE-----DSDSDSAEEDTAQTGATPHIT-SAPLKAWV 367  
Qy 487 YNPQNET-----ATIQTAARI---VPEEPDSSEKDLGSKGSDL-----E 521  
Db 368 YRGEUTETEENSOLDAAEEDTAQTGATPHITSAFLKAWVVRPGEDTEEE---NSDLSAAEE 424  
Qy 522 NSSQSGLPET-----BEHSGGEDDWEES-----SADAEASLKWNSFC 560  
Db 425 DTATGATGATPHITSAFLKAWVVRPGEDTEEDTEEESEENAVAPGSETADSSQSPCLQPC 484  
Qy 561 -----NSDDPYNPLNFKAPFOTSGENEGKCRDCKTSPSEIVALSECHTLLSCKVQLL 612  
Db 485 LPOEYKTKGGESE-PL-FQVAFYLPGEK-----PESWAAAPKPLRLQRLRLP 531  
Qy 613 GS-----QESCECPDSVQRDVLSGGRHTHVKRKVTFLEEVTEYYIS-----GDEDRKGPWE 663  
Db 532 KATRDQDPEIP-----LKAARKVHFAEKVTVHFLAVWAGPAQAARGPWE 576  
Qy 664 EPARDGCRFOKRIQIETEDAIYCLTFEHRERMFNRLQ 700  
Db 577 QFARDRSRFPARRIAQAEEKLGPYLTPTDSRARAWARLR 613

## RESULT 8

US-10-277-603-3  
; Sequence 3, Application US/10277603  
; Publication No. US20030219376A1  
; GENERAL INFORMATION:  
; APPLICANT: Fisher, Paul B.  
; TITLE OF INVENTION: PROGRESSION ELLEVATED GENE-3 AND USES  
; FILE REFERENCE: A34608-B-A  
; CURRENT APPLICATION NUMBER: US/10/277,603  
; CURRENT FILING DATE: 2002-10-22  
; PRIOR APPLICATION NUMBER: 09/052,753  
; PRIOR FILING DATE: 1998-03-31  
; PRIOR APPLICATION NUMBER: PCT/US98/05793  
; PRIOR FILING DATE: 1998-03-20  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 657  
; TYPE: PRT  
; ORGANISM: Rat  
US-10-277-603-3

Query Match 6.2%; Score 236; DB 15; Length 657;  
Best Local Similarity 21.8%; Pred. No. 1e-08;  
Matches 165; Conservative 77; Mismatches 225; Indels 290; Gaps 33;

Qy 70 LLAPLPGLLQKVLWISQLFGGMFPTRWLDPAQVYSALRALKGRKPAAPTAKSLSLQ 129  
Db 21 LLSPLMGLLSRA--WSRLRGPEVPEAWL----- 46  
Qy 130 DSDPSVTSPLDWLEEGIHQWYSPDLKLELKAGKSALDPAQAQAFLEQQLWGVLLPSS 189  
Db 47 ----AKTVTGADQIE-----AAALLTPTVSG-NLLP-- 73  
Qy 190 LQSLRYNRELGSPPSPLNIQRIIDNFVSVYLLNPSYLDGFPRLVSYQNSDGNSEVVG 249  
Db 74 ----HGTEESGSPESQQAQRL-----CLVEAESPPETWGLSNVDE 112  
Qy 250 FQTLTPSSCLRDHCHPQPLSAELIPASWQGCPLSTGLPEIHLHMKRLEFLQOANK 309  
Db 113 YNA-KPGQDRLREKEMERTAGKATLQAGLQAGD-----KRLGEVVAEE 156  
Qy 310 G----QDLPTPDQNGYHSLBEEHSLR-----MDPKHCRDNPOTQFVPAAGDIPGNTQ 358  
Db 157 GVAPAYPTSQLGGPAENEEDGETVTKYQASAASIAPGYKPTVPVFLGEA-----EH 210  
Qy 359 ESTEEKIELLTTEVPLALEBESSEGCPS--EIPMEKEPGEGRISVVDYSYLEGDLPI 416  
Db 211 QATEEK-----GTE-NKADPSNSPSSGSHRAWETYSREKPKQ-----EGEAKVE 254

QY 417 ARPA-----CSNKLID-----YILGASSDLETSDDPE----- 444  
DB 255 AHRAGQGHPCRNAEAEAGGPETTFVCTGNFLKAWVYRPGEDTEEDNSDSDSAEEDTAQ 314  
QY 445 -----GEOWDEAEADGFDSDSSLSLSDLEQDPEGLHLWNSFCSDVP 486  
DB 315 TGATPHTSAFLKAWVYRPGEDTEEB-----DSDSAEEDTAQTGATPHT-SAFLLKAW 367  
QY 487 YNPQNFT-----ATTQTAARI---VPEFSDSEKDLGSKSL---E 521  
DB 368 YRPGEDTEENSDDLSDAEEDTAQTGATPHTSAFLKAWVYRPGEDTEEB---NSOLDSABE 424  
QY 522 NSSQSGSLPET-----PEHSSGREDOWES-----SADAEASLKLWNSFC 560  
DB 425 DTAQTGATPHTSPFLKAWVYRPGEDTEDTTEEDSENVAQDSEADSSQSPCLQPORC 484  
QY 561 -----NSDDPNPLNFKAPFQTSGENEKGCRDSTPSSIVAISECHTLSSCKVOLL 612  
DB 485 LPGEKTKGRGEBP--PL-FQVAFYLPGEK-----PESPWAAAPKLPLRLQRLRLP 531  
QY 613 GS----QSECEPDSVQRDVLSGRHVVHKKVTFLEEVTEYIIS-----GDEDRKGWE 663  
DB 532 KAPTRDQDEIP-----LKAARKVHFAEKVTVHFLAVWAGPAQAARRGWE 576  
QY 664 EFARDGCRFKRIOETDAIGYCLTFEHRERMFNRQ 700  
DB 577 QFARDRSFARRIAQAEEKLGPYLTDPDSRAWAARLR 613

## RESULT 9

US-09-052-753-7  
; Sequence 7, Application US/09052753A  
; Patent No. US20010014734A1  
; GENERAL INFORMATION:

; APPLICANT: Fisher, Paul B.  
; TITLE OF INVENTION: PROGRESSION ELEVATED GENE-3 AND USES THEREOF  
; FILE REFERENCE: 0575/51523-B/JPW/JSG  
; CURRENT APPLICATION NUMBER: US/09/052,753A  
; CURRENT FILING DATE: 1998-03-31  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: Patentin ver. 2.1  
; SEQ ID NO 7

; LENGTH: 578  
; TYPE: PRT  
; ORGANISM: Human  
US-09-052-753-7

Query Match 5.8%; Score 219; DB 9; Length 578;  
Best Local Similarity 21.7%; Pred. No. 1.6e-07;  
Matches 155; Conservative 79; Mismatches 223; Indels 258; Gaps 34;

QY 53 SAOPETRVSYWTKLLS-QLLAPLPGLLQKVLVLSQLFGMFPTRWL--DFAGV----- 102  
DB 4 SPRPE-HVLHWKEAHSFYLLSPLMGFLSRA--WSRLRGPEVSEAWLAETVAGANQIQADA 60  
QY 103 -----YSALRAKLGREKPA-APTAQKSLSSLDSDSDPSVTSPLDWLEEGIHWOY 151  
DB 61 LLTPPPVSENHPLIRETEGNGTPPEWSKAAQRLCLDVEAQSSPPKT-----WGL 108  
QY 152 SPDDLKLELKAGSALDPAQAFLLEQOLMGVLLPSSLSQRLYSNRELSSPGPLNIQ 211  
DB 109 SDID---EHNGK-----PGQDGLREQVEHTAGLP-TLQ-----PLHLQ 143  
QY 212 RIDNFSVSYLLNPSYLDLCPFRLEVSQNSDGNSEVVGFTLTPESSCLREDHCHPQPLS 271  
DB 144 GADK-----KVGEVVAAREGVSELA----- 163  
QY 272 AELIPAS-WQGCP---PLSTEGPLPEIHHLMKRLFFLOQANKGQDLPTPDQDNGYHSLEE 327  
DB 164 ---YPTSHWEGPADEEDTETVKAH-----QASAASIAP-----GY----- 198  
QY 328 EHSLLRMDPKHCRDNPTQFVPAAGDI PGNTQESTEEKI-----ELLTTEVPLALE 377

DB 199 -----KPSTSVYCPGEAEHRAETBEKGTDNKAEPSSGSHSRFWFVHTRRPRKQEG 246  
QY 378 EESPSECCPSELP--MEKEPGEGRISVVDYSYLEGDLPLISARPACSNKLIDYILGGAS 434  
DB 247 ETKPEQHRAGQSHPCQNAESEEGPETS-----CS-----GS 279  
QY 435 SDLETSDPGEQDEWAEADGFDSDSSLSLSDLEQDPEGLHLWNSFCSDVPYNPQNFTA 494  
DB 280 AFLKAWVYRPGEDTEEB-EDPDLS-----AEED-----TAHTCTTPH-----TS 318  
QY 495 TIQTAAIRIVEEPESDSKDLGSKSDLENSSQSGSLPET-----PEHSSGEEDDWE 544  
DB 319 AFLKAWVYRPGEDTEEDDGDWDSAEADAQSCCTTPTHTSAFLKAWVYRPGEDTEEDDSE 378  
QY 545 SSAD-EAESLKLWNSFCNSDDPNPLNFKAPFQTSGENEKGCRDSTPSSIVAISECHT 603  
DB 379 NVAPVDSETV-----DSCQSTQHCLFV-----BKTGCGGEARPPPFQVA----- 417  
QY 604 LLSCKVOLLGSOES-----ECPDVSVQRDVLS-----GGRHVVHKKVTFLEEVTEY 650  
DB 418 -----FYLPGOKPAPWAAPKLPLRLQKRLRSFKAPARNQGPETPLKGRKVHSEKVTVH 472  
QY 651 YIS-----GDEDRKGWEFARDGCRFKRIOETDAIGYCLTFEHRERMFNRQ 700  
DB 473 FLAVWAGPAQAARRGPEWFARDRSFARRIAQAEEQLGPLYLTPAFARAWTRLR 527

## RESULT 10

US-10-277-603-7  
; Sequence 7, Application US/10277603  
; Publication No. US20030219376A1  
; GENERAL INFORMATION:

; APPLICANT: Fisher, Paul B.  
; TITLE OF INVENTION: PROGRESSION ELEVATED GENE-3 AND USES  
; TITLE OF INVENTION: THEREOF  
; FILE REFERENCE: A34608-B-A  
; CURRENT APPLICATION NUMBER: US/10/277,603  
; CURRENT FILING DATE: 2002-10-22  
; PRIOR APPLICATION NUMBER: 09/052,753  
; PRIOR FILING DATE: 1998-03-31  
; PRIOR APPLICATION NUMBER: PCT/US98/05793  
; PRIOR FILING DATE: 1998-03-20  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 7

; LENGTH: 578  
; TYPE: PRT  
; ORGANISM: Human  
US-10-277-603-7

Query Match 5.8%; Score 219; DB 15; Length 578;  
Best Local Similarity 21.7%; Pred. No. 1.6e-07;  
Matches 155; Conservative 79; Mismatches 223; Indels 258; Gaps 34;

QY 53 SAOPETRVSYWTKLLS-QLLAPLPGLLQKVLVLSQLFGMFPTRWL--DFAGV----- 102  
DB 4 SPRPE-HVLHWKEAHSFYLLSPLMGFLSRA--WSRLRGPEVSEAWLAETVAGANQIQADA 60  
QY 103 -----YSALRAKLGREKPA-APTAQKSLSSLDSDSDPSVTSPLDWLEEGIHWOY 151  
DB 61 LLTPPPVSENHPLIRETEGNGTPPEWSKAAQRLCLDVEAQSSPPKT-----WGL 108  
QY 152 SPDDLKLELKAGSALDPAQAFLLEQOLMGVLLPSSLSQRLYSNRELSSPGPLNIQ 211  
DB 109 SDID---EHNGK-----PGQDGLREQVEHTAGLP-TLQ-----PLHLQ 143  
QY 212 RIDNFSVSYLLNPSYLDLCPFRLEVSQNSDGNSEVVGFTLTPESSCLREDHCHPQPLS 271  
DB 144 GADK-----KVGEVVAAREGVSELA----- 163  
QY 272 AELIPAS-WQGCP---PLSTEGPLPEIHHLMKRLFFLOQANKGQDLPTPDQDNGYHSLEE 327

Db 164 ---YPTSHWGGPAEDDEETETVKAH-----QASAAIAP-----GY----- 198  
 Qy 328 EHSLLRMDPKHCRDNPTQFVPAAGDIPGNTQESTEEKI-----ELLTTTEVPLALE 377  
 Db 199 -----KPTSVYCPGEAEHRAATEEKGTDNKAEPGSHSRFMEYHTRERPKQEG 246  
 Qy 378 BEPSBECPCSEIP---MEKEPBGGRISVVDYSLYEGDLPI SARPACSNKLIDYILGGAS 434  
 Db 247 ETKPEQHRAGQSHPCQNAESEEGPETS-----CS-----GS 279  
 Qy 435 SDETSDDPGEOWDEAEADGDSLSLSDLEQDPEGLHWNFSVDPYNPQNFTA 494  
 Db 280 AFLKAWYRPGEDTEEB-EDPDLS-----AEED-----TAHTCTTPH-----TS 318  
 Qy 495 TIQTAARIVPEEPSDSKOLSGKSDLENSQSGSLPET-----PEHSGEEDDWE 544  
 Db 319 AFLKAWYRPGEDTEEDDDGWDSDAEADAQSCITPHTSAFLKAWYRPGEDTEEDDSE 378  
 Qy 545 SSAD-EAESLKLWNSFCNSDDPNYNLFKAPFQTSGENEKGCRDSTPSSSIVAISECHT 603  
 Db 379 NVAPVDSETV---DSCQSTQHCLFV-----EKTGCGEAEPPFPQVA----- 417  
 Qy 604 LLSCKVQLLSQES-----ECPDSVQDVLS-----GGRHTHVKRKVTPLFEEVTEY 650  
 Db 418 -----FYLPOQKAPPWAAPKLPLRLQRLRSFKAPARNQGPETPLGRKRVHSEKVTVH 472  
 Qy 651 YIS-----GBEDRKGWEEFARDGCRFQKRIQETEDAIGVCLTFEHRMFMNRLQ 700  
 Db 473 FLAWAGPAQAARGPWEQLARDRSRFAIRIAQAEKJGPLYLTPAFARAWARL 527

RESULT 11

US-09-552-753-2  
 ; Sequence 2, Application US/09052753A  
 ; Patent No. US20010014734A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Fisher, Paul B.  
 ; TITLE OF INVENTION: PROGRESSION ELEVATED GENE-3 AND USES THEREOF  
 ; FILE REFERENCE: 0575/51523-B/JPW/JSG  
 ; CURRENT APPLICATION NUMBER: US/09/052,753A  
 ; CURRENT FILING DATE: 1998-03-31  
 ; NUMBER OF SEQ ID NOS: 8  
 ; SOFTWARE: Patent In Ver. 2.1  
 ; SEQ ID NO 2  
 ; TYPE: PRT  
 ; LENGTH: 590  
 ; ORGANISM: Rat  
 ; US-09-552-753-2

Query Match 5.3%; Score 201; DB 9; Length 590;  
 Best Local Similarity 23.9%; Pred. No. 3.8e-06;  
 Matches 142; Conservative 56; Mismatches 210; Indels 186; Gaps 29;  
 Qy 245 SEVVGFQTLTPESSCLREDHCHPOPLSAEL-----IPASWQCCPPLSTEGLPRIHILRM 298  
 Db 23 SPLMGF--LSRAWSRLRVPEA-PEPWAETVTGADQIADAHAPAPPL-----VPE-NHPPO 74  
 Qy 299 KRLEFLQANKGQDLPTPDQDNGVHSLREHSHLLRMDPKHCRDNP-----TQFVP--- 348  
 Db 75 GEABESGTPBEGKAAQGPCLDVQANSPPETGLSDDDKQGGQPGQRAHTAGLPILL 134  
 Qy 349 -----AAGDIPGNTQESTEEKIELLTTEVPLALEEESPSCGCPSS- 389  
 Db 135 SPGLQADKSLGEVVAEGVTE-----LAV-PTSHWEGCPSEEEDEGTVKKA 182  
 Qy 390 -----IPMEKE-----PBGGRISVVDYSLYEGDLPI SARPACSNKLIDYILGGASDL 437  
 Db 183 FRASADSPGHKSSTSVYCPGEAEHQATEEKQTKENKADPPSSPS-----GSHRAW 232  
 Qy 438 ETSDDPGEOWDE-----EAEDDGFDSLSLSDLEQDPEGLHWNFSV 484  
 Db 233 EYCSKQEGADPEPHRAGKYQLCONAEAEAEAEAEKVSLSVSS-----GNAFLKA 282

Qy 485 DPYNPQNT-----ATIQTAAARIVPEEP-----SDSEKOLSGKSDLE 521  
 Db 283 WYVRPGEDTDDDDSDWGSAAEEGKALSPTSPEHDFLKAWYRPGEDTDD--DDSDWG 340  
 Qy 522 NSSQSG---SLPPTPEHS-----SGE--EDDWESSADEAES-----L 553  
 Db 341 SAESEKALSSPTSPEHDFLKAWYRPGEDTDDQSDWGSAAEKDGLAQTFATPHTSAPL 400  
 Qy 554 KLWNSFCNSDDPNY-----PLNFKA--PFQTSGENEKGCRDSTPSSSIVAISECHT 603  
 Db 401 KTW-VCCPGEDTDDCEVVVVPEDSEADPDKSPSHEAQGC-----LPGEQTEGLVAEHS 455  
 Qy 604 LLSCKVQLLSQES-----ECPDSVQDVLS-----GGRHTHVKRKVTPLFEEVTEY 650  
 Db 456 LFOVAFYLFCEKPAFPWPTAPKLPLRLQRLTLRLTPTQDDPETPLARAKVHSEKVTVH 515  
 Qy 651 YIS-----GBEDRKGWEEFARDGCRFQKRIQETEDAIGVCLTFEHRMFMNRL 699  
 Db 516 FLAWAGPAQAARGPWEQLARDRSRFAIRIAQAEKJGPLYLTPAFARAWARL 569

RESULT 12

US-10-277-603-2  
 ; Sequence 2, Application US/10277603  
 ; Publication No. US20030219376A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Fisher, Paul B.  
 ; TITLE OF INVENTION: PROGRESSION ELEVATED GENE-3 AND USES  
 ; FILE REFERENCE: A34608-B-A  
 ; CURRENT APPLICATION NUMBER: US/10/277,603  
 ; CURRENT FILING DATE: 2002-10-22  
 ; PRIOR APPLICATION NUMBER: 09/052,753  
 ; PRIOR FILING DATE: 1998-03-31  
 ; PRIOR APPLICATION NUMBER: PCT/US98/05793  
 ; PRIOR FILING DATE: 1998-03-20  
 ; NUMBER OF SEQ ID NOS: 8  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 2  
 ; TYPE: PRT  
 ; LENGTH: 590  
 ; ORGANISM: Rat  
 ; US-10-277-603-2

Query Match 5.3%; Score 201; DB 15; Length 590;  
 Best Local Similarity 23.9%; Pred. No. 3.8e-06;  
 Matches 142; Conservative 56; Mismatches 210; Indels 186; Gaps 29;  
 Qy 245 SEVVGFQTLTPESSCLREDHCHPOPLSAEL-----IPASWQCCPPLSTEGLPRIHILRM 298  
 Db 23 SPLMGF--LSRAWSRLRVPEA-PEPWAETVTGADQIADAHAPAPPL-----VPE-NHPPO 74  
 Qy 299 KRLEFLQANKGQDLPTPDQDNGVHSLREHSHLLRMDPKHCRDNP-----TQFVP--- 348  
 Db 75 GEABESGTPBEGKAAQGPCLDVQANSPPETGLSDDDKQGGQPGQRAHTAGLPILL 134  
 Qy 349 -----AAGDIPGNTQESTEEKIELLTTEVPLALEEESPSCGCPSS- 389  
 Db 135 SPGLQADKSLGEVVAEGVTE-----LAV-PTSHWEGCPSEEEDEGTVKKA 182  
 Qy 390 -----IPMEKE-----PBGGRISVVDYSLYEGDLPI SARPACSNKLIDYILGGASDL 437  
 Db 183 FRASADSPGHKSSTSVYCPGEAEHQATEEKQTKENKADPPSSPS-----GSHRAW 232  
 Qy 438 ETSDDPGEOWDE-----EAEDDGFDSLSLSDLEQDPEGLHWNFSV 484  
 Db 233 EYCSKQEGADPEPHRAGKYQLCONAEAEAEAEAEKVSLSVSS-----GNAFLKA 282  
 Qy 485 DPYNPQNT-----ATIQTAAARIVPEEP-----SDSEKOLSGKSDLE 521  
 Db 283 WYVRPGEDTDDDDSDWGSAAEEGKALSPTSPEHDFLKAWYRPGEDTDD--DDSDWG 340  
 Qy 522 NSSQSG---SLPPTPEHS-----SGE--EDDWESSADEAES-----L 553

Db 341 SAEEGKALSPTSFEHDFLKAWVYRPGEDTEDDQSDWGAEXDGLAQATPATHTSAFL 400  
Qy 554 KWNISFNCSDPDYN-----PLMFKA--PQTSGENEKGCRDCKTTPSESIVALSSEC-HT 603  
Db 401 KTW-VCCPGEDEDDCEVVVPEIDSEAADPKSPSHEAQC-----LPQEQTEGLVEAHS 455  
Qy 604 LLSCKVQLLGSQES-----ECPDVSQVDVL-----SGGRHTHVRKKVTELEEVTEY 650  
Db 456 LFQVAFYLPGEKPPWTAAPLPLRLQRLTLRLTPTQDQDPETPLRKARKVHFSENVTVH 515  
Qy 651 YIS-----GDEDRKGPWEFARDGCRFQKRIQETEDAIQCYCLTFEHRERMFNRL 699  
Db 516 FLAVWAGPAQAARRGFWQLARDRSFARRIAQAEEKLGPYLTFAFRARAWARL 569

## RESULT 13

US-09-902-432-4  
; Sequence 4, Application US/09902432  
; Patent No. US2002016002A1  
; GENERAL INFORMATION:  
; APPLICANT: Irwin H. Gelman  
; TITLE OF INVENTION: TUMOR SUPPRESSOR GENE  
; FILE REFERENCE: A30558-A-FWC-A 070156.0597  
; CURRENT APPLICATION NUMBER: US/09/902,432  
; CURRENT FILING DATE: 2002-04-08  
; PRIOR APPLICATION NUMBER: 08/978,277  
; PRIOR FILING DATE: 1997-11-25  
; PRIOR APPLICATION NUMBER: 08/665,401  
; PRIOR FILING DATE: 1996-06-18  
; PRIOR APPLICATION NUMBER: 08/635,121  
; PRIOR FILING DATE: 1996-04-19  
; NUMBER OF SEQ ID NOS: 35  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 4  
; LENGTH: 1596  
; TYPE: PRT  
; ORGANISM: Rattus norvegicus  
US-09-902-432-4

Query Match 4.4%; Score 167; DB 9; Length 1596;

Best Local Similarity 19.8%; Pred. No. 0.0061;

Matches 151; Conservative 112; Mismatches 286; Indels 214; Gaps 35;

Qy 6 GGSRKRL-----GPR-----AGFRFWPPPPRRSQAGSSKEPT 38  
Db 673 GSSKRRARKASSDDDEGGFRTLGDSHRAEASKDKEAGTDAVPASTQEQDAQGSSPE 732  
Qy 39 PLGPENSGNPTLLSSAQPETRVSYWTKLLSQLLAPLGLQKV-----LIWSQLFGGMF 92  
Db 733 PAGSPSEGE-----GVSTW-ESFKLVTPRKSKSLBEKAEDSSVEQLSTEIE 780  
Qy 93 PTR---WLDFAGVYSALRAKREKPAAPTAKSLS-----SLQDSSDPVTS--PLDW 142  
Db 781 PSREESWV-----SIKKFIPGRKRRADGQEQATVDSGTVNEIDDPNVPVAVPLS- 833  
Qy 143 LEEGHHWQYSPDLKLELKAGSALDPAQAFLLEQLQWGVLLPSSLSQSLYSLNRELS 202  
Db 834 -----EYNAVE-REKMEAGNTPEL-----QLLGA-----VYVSEELSK 866  
Qy 203 SPSPGLNQRIDNFSVSYL--LNPSYLDCE-----PRLEVSQNSDGNSE 246  
Db 867 TLVHTVSVAVIDGTNAVTSVERSPSWISASVTEPLEHTAGEAMPVVEVTEKIDIAET 926  
Qy 247 VVGFTQLPSSCLREHCHPO-PLSABELIPASWQCGCPPLSTEGPEIHHHMKRLBFLQ 305  
Db 927 PVLQTL-PEGKADHDDMVTSEVDFTSEAVTAT-ETSEALTEEVTEASGAE-ETTDVVS 983  
Qy 306 QANKQDLP-----TPDQ--DNGVHSLSEE-----HSLLRMDPKCRDN-----PTQFVPA 349  
Db 984 AVSQTDSPTTEATPQVEESGVLDTEEBERQTAQLQAVDAVKVEESQVPATQTVQR 1043

Qy 350 AGDIPGNTQSTBEKIELLTTEVPLALEEBSPGECPSSEIPMEKEGEGRISVVVDYSYL 409  
Db 1044 TGSKALEKVEVEEDSEVLASEKEKDVPKGPVQEAHAHQSGSETGQATPESLEVPEV 1103  
Qy 410 EGDLPISARPACS-----NKLIDYILGAGSADLTFTSSDPEG-----EDWDEEAEDDGF 458  
Db 1104 TADVHVHA--TCQVIKLQQLMEQAAPESSETLTDSETNGSTPLADSDTADGTQDQDETID 1161  
Qy 459 SDSLSDSLLEQDPEGLHWNFSCSDVPYNPQNFATIQTAARIVPREPS-----D 509  
Db 1162 SQDSKATAAARQS-----QVTEBEAATAQKEPSTLPNNVPAQE 1200  
Qy 510 SEKDLGSKDLENSQ---SGSLPETPEHSGBED--DW---ESSADEASLKLWNSFCN 561  
Db 1201 EHGEPEGRDVLPTQOQELTAAAVPLAKTEVGQEGEVDWLDGEKVKEQE----- 1250  
Qy 562 SDDPYNPLNFKAPQTSGENEKGCRGSKTSPESIVALSECHTLLSKVQLLGSQESCPD 621  
Db 1251 -----VFVHSGPNSSQKAADVTDSE--VMGVAGCQKEKSTEVQSLSEEGEMET 1297  
Qy 622 SVQRDVLSSGGRHTHVRKKVTFLEEVTEYYISGDEDRKGPWEE 664  
Db 1298 DVE-----KEKRETKPEQVSE---EGEQETAPEHE 1325

## RESULT 14

US-10-097-534-13  
; Sequence 13, Application US/10097534  
; Publication No. US20030049607A1  
; GENERAL INFORMATION:  
; APPLICANT: GREENER, TSVIKA  
; APPLICANT: MOSKOWITZ, HAIM  
; APPLICANT: REISS, YUVAL  
; APPLICANT: ALROY, IRIS  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE MODULATION OF VIRAL  
; FILE REFERENCE: PLV-001.01  
; CURRENT APPLICATION NUMBER: US/10/097,534  
; PRIOR FILING DATE: 2002-03-12  
; PRIOR APPLICATION NUMBER: 60/275,224  
; PRIOR FILING DATE: 2001-03-12  
; PRIOR APPLICATION NUMBER: 60/308,958  
; PRIOR FILING DATE: 2001-07-31  
; PRIOR APPLICATION NUMBER: 60/340,170  
; PRIOR FILING DATE: 2001-12-07  
; NUMBER OF SEQ ID NOS: 71  
; SOFTWARE: PatentIn ver. 2.1  
; SEQ ID NO 13  
; LENGTH: 1562  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-097-534-13

Query Match 4.2%; Score 160.5; DB 14; Length 1562;

Best Local Similarity 20.9%; Pred. No. 0.018;

Matches 161; Conservative 99; Mismatches 277; Indels 235; Gaps 40;

Qy 31 AGSKKPTTP-LGPNESGNPTLLSSAQPETRVSYWT-KLLSQLLAPLPGL-LQKVLWSQL 87  
Db 123 SGALRATTPSVTVKNSAAPFKSIGADETVQGGSRRLISFSLDSDFQAMGLKK----- 175  
Qy 88 FGGMF--PTRWLQFA---GVYSALRAL--KGREKPAAPTAKQSLSSQLDSDSPSVTSPL 140  
Db 176 --GMFNPDPYKLISIQPGKHSIFPALPHGQER-----RSKLI-----NTVNP 219  
Qy 141 DWLEEGHWWQYSPDP-LKLELKAGSALDPAQAFA-----LLEQQLWGVLLPSS 189  
Db 220 -WQAQEQSFVSLPTDVLIEIVKDKFAKSRPIIKRFLGKLSMPVQRLIERHAIGDRVVSYT 278  
Qy 190 LQRLYSNRELG-----SSPSGPLNIQRIDNFSVSYLLNPSYLD 230  
Db 279 LGRRLPDTHVSGQLQFRFEITSSIHDPDEISISTEPESAQIQD--SPMNLMSGSGE- 335

```
Qy 231 PPRLEVSQNSDGNSEVVGFTLTPESSCLREDHCHPQPLSAELIPASWGCPPLSTEG 290
Db 336 -PREASESSEWKEQLEGSDRPNQNSIELSRPAEEAAVITEAGDQGMVSVGPEGA 394
Qy 291 PEIHLRMKLEFLQANKG-QDLPTDQDNGHSLSEEHSLLRMDPKHCRDNPQTQVPA 349
Db 395 GEL-----LAQVKDIQAPSAEALAEOLDLGEASALLD----- 431
Qy 350 AGDIPGNTQESTKEIKLTTEVPLALEE---ESPSEGCPSSEIPMEKEPGEGRISVVYD 406
Db 432 -GEAPASTKEEPL- -BEATTQSRAGREBEKEQEEGDVST-----LEQGEGLQ---- 479
Qy 407 SYLEGDLPIGARPACSNKL-IDVILGASDLET-----SSDPEGEDWDEEA 452
Db 480 --LRASVYKRSRP-CSLPVSELETVIASACGDPETPRTHYIRHTLHSMPSAQGGSAE 536
Qy 453 EDGCFDSDSLSDLEQDPEGLHWNPFCSVDYPNQNTATIQTAARIVPEPSPSEK 512
Db 537 EEDGAEESTLKDS---SEKQGL-----SEVDTV-----ADPSALEE 571
Qy 513 DLGKSDLENSSQSGSLPET--PEHSSG-----BEDDWESSADEAESLKLWNSF 559
Db 572 D-----REEPEGATPGTAHPGSHGHPPLSANGAQQDTHPTGSESDS----- 616
Qy 560 CNSDDPYNPLNFKAPFTSGENEKGRDSTKTPSESIVAISECHTLTLLSCKVQLGSESEC 619
Db 617 -----SPRQGGDHSCGEC-DASCCSPSCYS-SSCYS-TSCYSSSCYSASCYS 660
Qy 620 PDSVQRDVLSGGR---HTHVKRKVTFLVEVTEYIISGDEDRK----- 659
Db 661 P-----SCYNGNRFASHTRFSSVDSAKISESTVFSSQDDSEENSAPESVPSMQSPELD 715
Qy 660 -----GPW-BEFARDGCRFQKRIQETDAIGYCLTFEHRERMFNRLOQTC 703
Db 716 PESTNGAGPWQDELAAPSGHVERSPGLESPVA-----GPSNRREGEC 758

RESULT 15
US-10-450-763-57852
; Sequence 57852, Application US/10450763
; Publication No. US20050196754A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 790CIP3/US
; CURRENT APPLICATION NUMBER: US/10/450,763
; CURRENT FILING DATE: 2003-06-11
; PRIOR APPLICATION NUMBER: PCT/US01/08631
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/649,167
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 60736
; SOFTWARE: Custom
; SEQ ID NO 57852
; LENGTH: 1562
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (1524)..(1556)
; OTHER INFORMATION: HECT-domain (ubiquitin-transferase) domain identified by
; OTHER INFORMATION: EMATRIX, accession number PF00632C, p-values=3.302e-23, raw score
; OTHER INFORMATION: 20.66
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (1257)..(1562)
; OTHER INFORMATION: HECT-domain (ubiquitin-transferase) domain identified by
; OTHER INFORMATION: Pfam, accession name HECT, E-value=4.3e-127, Pfam score of 435.6
US-10-450-763-57852
```

Query Match 4.2%; Score 160.5; DB 18; Length 1562;

```
Best Local Similarity 20.9%; Pred. No. 0.018;
Matches 161; Conservative 99; Mismatches 277; Indels 235; Gaps 40;
Qy 31 AGSSKEPTP-LGPENSGNPTLLSSAQPETRVSYWT-KLLSOLLAPLPGL-LQKVLWSQL 87
Db 123 SGALRTTSVTWNSAAPIFKISIGADETVQGGSRRLISFSLDFQMLGLK----- 175
Qy 88 FGMF--PTRWLDF--GVYSALRAL--KGREKPAAPTAQKLSLSLQDSSDPSVTSPL 140
Db 176 --GMFPNDPPLYKISIQPGKHSIFPALPHHQBQER-----RSKIIG-----NTWNP 219
Qy 141 DWLEEGHWQYSPD-LKLELKAKGSALDPAQAFA-----LLEQQLWGVLLPSS 189
Db 220 -WQAEQFSFVSLPTDLEITEVDKFAKSRPIIKRFLGKLSMPVQRLERHAIGDRVVSYT 278
Qy 190 LQRLYSNRELG-----SSPSGPNLQIRIDNFSVSVLLNPSVLD 230
Db 279 LGRRLTDHVGQQLQRFPEITSIHPDDDEISLSTEPESAQIQD--SPMNLMESSGE- 335
Qy 231 FPRLEVSQNSDGNSEVVGFTLTPESSCLREDHCHPQPLSAELIPASWGCPPLSTEG 290
Db 336 -PREASESSEWKEQLEGSDRPNQNSIELSRPAEEAAVITEAGDQGMVSVGPEGA 394
Qy 291 PEIHLRMKLEFLQANKG-QDLPTDQDNGHSLSEEHSLLRMDPKHCRDNPQTQVPA 349
Db 395 GEL-----LAQVKDIQAPSAEALAEOLDLGEASALLD----- 431
Qy 350 AGDIPGNTQESTKEIKLTTEVPLALEE---ESPSEGCPSSEIPMEKEPGEGRISVVYD 406
Db 432 -GEAPASTKEEPL- -BEATTQSRAGREBEKEQEEGDVST-----LEQGEGLQ---- 479
Qy 407 SYLEGDLPIGARPACSNKL-IDVILGASDLET-----SSDPEGEDWDEEA 452
Db 480 --LRASVYKRSRP-CSLPVSELETVIASACGDPETPRTHYIRHTLHSMPSAQGGSAE 536
Qy 453 EDGCFDSDSLSDLEQDPEGLHWNPFCSVDYPNQNTATIQTAARIVPEPSPSEK 512
Db 537 EEDGAEESTLKDS---SEKQGL-----SEVDTV-----ADPSALEE 571
Qy 513 DLGKSDLENSSQSGSLPET--PEHSSG-----BEDDWESSADEAESLKLWNSF 559
Db 572 D-----REEPEGATPGTAHPGSHGHPPLSANGAQQDTHPTGSESDS----- 616
Qy 560 CNSDDPYNPLNFKAPFTSGENEKGRDSTKTPSESIVAISECHTLTLLSCKVQLGSESEC 619
Db 617 -----SPRQGGDHSCGEC-DASCCSPSCYS-SSCYS-TSCYSSSCYSASCYS 660
Qy 620 PDSVQRDVLSGGR---HTHVKRKVTFLVEVTEYIISGDEDRK----- 659
Db 661 P-----SCYNGNRFASHTRFSSVDSAKISESTVFSSQDDSEENSAPESVPSMQSPELD 715
Qy 660 -----GPW-BEFARDGCRFQKRIQETDAIGYCLTFEHRERMFNRLOQTC 703
Db 716 PESTNGAGPWQDELAAPSGHVERSPGLESPVA-----GPSNRREGEC 758
```

Search completed: September 16, 2005, 11:04:29  
Job time : 245.057 secs

**THIS PAGE BLANK (USPTO,**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 16, 2005, 10:40:09 ; Search time 182.419 Seconds  
(without alignments)  
4565.254 Million cell updates/sec

Title: US-10-650-482-2

Perfect score: 3804

Sequence: 1 MEFGTGGRKRLGPRAGFRF.....RMFNRLLQGTCTFKGLNVLKQC 713

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 6959266 seqs, 1168006243 residues

Total number of hits satisfying chosen parameters: 6959266

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Databases :

Pending\_Patents\_AA\_Main:\*  
1: /cgn2\_6/ptodata/1/paa/US06 COMB.pcp.\*  
2: /cgn2\_6/ptodata/1/paa/US06 COMB.pcp.\*  
3: /cgn2\_6/ptodata/1/paa/US07 COMB.pcp.\*  
4: /cgn2\_6/ptodata/1/paa/US08 COMB.pcp.\*  
5: /cgn2\_6/ptodata/1/paa/US08 COMB.pcp.\*  
6: /cgn2\_6/ptodata/1/paa/US08 COMB.pcp.\*  
7: /cgn2\_6/ptodata/1/paa/US08 COMB.pcp.\*  
8: /cgn2\_6/ptodata/1/paa/US08 COMB.pcp.\*  
9: /cgn2\_6/ptodata/1/paa/US08 COMB.pcp.\*  
10: /cgn2\_6/ptodata/1/paa/US08 COMB.pcp.\*  
11: /cgn2\_6/ptodata/1/paa/US08 COMB.pcp.\*  
12: /cgn2\_6/ptodata/1/paa/US08 COMB.pcp.\*  
13: /cgn2\_6/ptodata/1/paa/US08 COMB.pcp.\*  
14: /cgn2\_6/ptodata/1/paa/US09 COMB.pcp.\*  
15: /cgn2\_6/ptodata/1/paa/US09 COMB.pcp.\*  
16: /cgn2\_6/ptodata/1/paa/US09 COMB.pcp.\*  
17: /cgn2\_6/ptodata/1/paa/US09 COMB.pcp.\*  
18: /cgn2\_6/ptodata/1/paa/US09 COMB.pcp.\*  
19: /cgn2\_6/ptodata/1/paa/US09 COMB.pcp.\*  
20: /cgn2\_6/ptodata/1/paa/US09 COMB.pcp.\*  
21: /cgn2\_6/ptodata/1/paa/US09 COMB.pcp.\*  
22: /cgn2\_6/ptodata/1/paa/US09 COMB.pcp.\*  
23: /cgn2\_6/ptodata/1/paa/US09 COMB.pcp.\*  
24: /cgn2\_6/ptodata/1/paa/US09 COMB.pcp.\*  
25: /cgn2\_6/ptodata/1/paa/US09 COMB.pcp.\*  
26: /cgn2\_6/ptodata/1/paa/US09 COMB.pcp.\*  
27: /cgn2\_6/ptodata/1/paa/US09 COMB.pcp.\*  
28: /cgn2\_6/ptodata/1/paa/US09 COMB.pcp.\*  
29: /cgn2\_6/ptodata/1/paa/US09 COMB.pcp.\*  
30: /cgn2\_6/ptodata/1/paa/US09 COMB.pcp.\*  
31: /cgn2\_6/ptodata/1/paa/US09 COMB.pcp.\*  
32: /cgn2\_6/ptodata/1/paa/US09 COMB.pcp.\*  
33: /cgn2\_6/ptodata/1/paa/US09 COMB.pcp.\*  
34: /cgn2\_6/ptodata/1/paa/US09 COMB.pcp.\*  
35: /cgn2\_6/ptodata/1/paa/US09 COMB.pcp.\*  
36: /cgn2\_6/ptodata/1/paa/US09 COMB.pcp.\*  
37: /cgn2\_6/ptodata/1/paa/US09 COMB.pcp.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	3804	100.0	713	32	US-10-650-482-2	Sequence 2, Appl1
2	3799	99.9	713	27	US-10-170-205E-23916	Sequence 23916, A
3	3799	99.9	713	37	US-60-452-680-17655	Sequence 17655, A
4	3794	99.7	713	20	US-09-629-469A-18965	Sequence 18965, A
5	3794	99.7	713	35	US-10-917-503-18965	Sequence 18965, A
6	3789	99.6	720	22	US-09-757-028-2348	Sequence 2348, Ap
7	3789	99.6	720	28	US-10-222-911-2348	Sequence 2348, Ap
8	3627	95.3	707	1	PCT-US00-05883-941	Sequence 941, App
9	3627	95.3	707	24	US-09-925-299-941	Sequence 941, App
10	2223	58.4	698	32	US-10-650-482-4	Sequence 4, Appl1
11	813	21.4	153	1	PCT-US00-35017A-963	Sequence 963, App
12	813	21.4	153	28	US-10-296-115-963	Sequence 963, App
13	511	13.4	101	1	PCT-US01-14827-15743	Sequence 15743, A
14	402	10.6	73	37	US-60-177-571-4491	Sequence 4491, Ap
15	399	10.5	73	37	US-60-162-247-4518	Sequence 4518, Ap
16	399	10.5	73	37	US-60-169-840-8186	Sequence 8186, Ap
17	334	8.8	67	37	US-60-196-718-7356	Sequence 7356, Ap
18	263	6.9	697	34	US-10-821-234-905	Sequence 905, App
19	260.5	6.8	674	27	US-10-170-205E-18860	Sequence 18860, A
20	260.5	6.8	674	28	US-10-219-051B-549	Sequence 549, App
21	260.5	6.8	674	28	US-10-219-051B-553	Sequence 553, App
22	260.5	6.8	674	28	US-10-219-051B-10552	Sequence 10552, A
23	260.5	6.8	674	28	US-60-443-566-4148	Sequence 4148, Ap
24	260.5	6.8	674	37	US-60-452-680-23269	Sequence 23269, A
25	260.5	6.8	674	37	US-60-453-050-14551	Sequence 14551, A
26	260.5	6.8	674	37	US-60-453-135-14551	Sequence 14551, A
27	260.5	6.8	674	37	US-60-453-444-7838	Sequence 7838, Ap
28	260.5	6.8	674	37	US-60-466-412-14551	Sequence 14551, A
29	260.5	6.8	674	37	US-60-466-412-14551	Sequence 14551, A
30	259.5	6.8	674	20	US-09-629-469A-11494	Sequence 11494, A
31	259.5	6.8	674	33	US-10-755-889-90	Sequence 90, Appl
32	259.5	6.8	674	35	US-10-917-503-11494	Sequence 11494, A
33	259.5	6.8	674	37	US-60-440-068-90	Sequence 90, Appl
34	259.5	6.8	674	37	US-60-469-757-90	Sequence 90, Appl
35	252.5	6.6	674	12	US-08-893-852-1	Sequence 1, Appl1
36	237	6.2	657	12	US-08-893-852-3	Sequence 3, Appl1
37	236	6.2	657	28	US-10-274-666-3	Sequence 3, Appl1
38	236	6.2	657	28	US-10-277-603-3	Sequence 3, Appl1
39	219	5.8	578	16	US-09-236-772-2	Sequence 2, Appl1
40	219	5.8	578	28	US-10-274-666-7	Sequence 7, Appl1
41	219	5.8	578	28	US-10-277-603-7	Sequence 7, Appl1
42	218.5	5.7	627	21	US-09-724-676-55843	Sequence 55843, A
43	218.5	5.7	627	21	US-09-724-676A-55843	Sequence 55843, A
44	201	5.3	590	12	US-08-893-852-4	Sequence 4, Appl1
45	201	5.3	590	28	US-10-274-666-2	Sequence 2, Appl1
46	201	5.3	590	28	US-10-277-603-2	Sequence 2, Appl1
47	170	4.5	1095	1	PCT-US03-38896-184	Sequence 184, App
48	170	4.5	1095	27	US-10-170-205E-22541	Sequence 22541, A
49	170	4.5	1095	37	US-60-452-680-14318	Sequence 14318, A
50	170	4.5	1095	37	US-60-453-050-8816	Sequence 8816, Ap
51	170	4.5	1095	37	US-60-453-135-8816	Sequence 8816, Ap
52	170	4.5	1095	37	US-60-466-412-8816	Sequence 8816, Ap
53	168	4.4	1596	1	PCT-US03-28227-5026	Sequence 5026, Ap
54	167	4.4	1596	1	PCT-US97-06830-4	Sequence 4, Appl1
55	167	4.4	1596	10	US-08-665-401-4	Sequence 4, Appl1
56	167	4.4	1596	24	US-09-902-432-4	Sequence 4, Appl1
57	167	4.4	1596	28	US-10-219-051B-8255	Sequence 8255, Ap
58	160.5	4.2	1562	1	PCT-US01-08631-57852	Sequence 57852, A
59	160.5	4.2	1562	26	US-10-097-534-13	Sequence 13, Appl1
60	160.5	4.2	1585	30	US-10-487-132-1	Sequence 1, Appl1
61	160.5	4.2	1607	18	US-09-488-725A-2337	Sequence 2337, Ap
62	160.5	4.2	1607	28	US-10-258-898A-2337	Sequence 2337, Ap
63	160.5	4.2	1664	18	US-10-286-897-2337	Sequence 2337, Ap
64	160.5	4.2	1664	18	US-09-488-725A-5909	Sequence 5909, Ap
65	160.5	4.2	1664	28	US-10-258-898A-5909	Sequence 5909, Ap
66	160.5	4.2	1664	28	US-10-286-897-5909	Sequence 5909, Ap
67	160.5	4.2	1702	1	PCT-US01-08631-38864	Sequence 38864, A
68	160	4.2	5303	37	US-60-173-464-23675	Sequence 23675, A
69	159	4.2	890	1	PCT-US01-14827-12944	Sequence 12944, A



70 158.5 4.2 1585 35 US-10-990-328-9765 Sequence 9765, Ap  
71 158.5 4.2 4866 37 US-60-191-637-29941 Sequence 29941, A  
72 158.5 4.2 4866 37 US-60-191-681-24058 Sequence 24058, A  
73 158.5 4.2 5303 20 US-09-614-150-30390 Sequence 30390, A  
74 158.5 4.2 5303 20 US-09-614-150A-30390 Sequence 30390, A  
75 158 4.2 2768 20 US-09-614-150-31983 Sequence 31983, A  
76 158 4.2 2768 20 US-09-614-150A-31983 Sequence 31983, A  
77 158 4.2 2768 37 US-60-173-464-24710 Sequence 24710, A  
78 158 4.2 2768 37 US-60-191-637-31555 Sequence 31555, A  
79 158 4.2 2768 37 US-60-191-681-25036 Sequence 25036, A  
80 157 4.1 866 27 US-10-170-205E-7942 Sequence 7942, Ap  
81 157 4.1 866 37 US-60-455-444-7969 Sequence 7969, Ap  
82 157 4.1 866 37 US-60-465-241-7969 Sequence 7969, Ap  
83 157 4.1 897 1 PCT-US03-14450-45 Sequence 45, Appl  
84 157 4.1 1772 30 US-10-437-963-143280 Sequence 143280, A  
85 157 4.1 2766 25 US-09-964-956-62 Sequence 62, Appl  
86 157 4.1 2766 31 US-10-514-150-6 Sequence 6, Appl  
87 156.5 4.1 1454 27 US-10-170-205E-20817 Sequence 20817, A  
88 156.5 4.1 1503 20 US-09-649-996-14 Sequence 14, Appl  
89 156 4.1 888 1 PCT-US01-08631-44262 Sequence 44262, A  
90 156 4.1 888 1 PCT-US01-08631-54269 Sequence 54269, A  
91 156 4.1 888 1 PCT-US04-07412-1795 Sequence 1795, Ap  
92 156 4.1 888 29 US-10-389-559-1795 Sequence 603, App  
93 156 4.1 888 37 US-60-365-264-603 Sequence 17202, A  
94 155.5 4.1 1801 30 US-10-438-246-17202 Sequence 6540, Ap  
95 155 4.1 440 27 US-10-179-131-6540 Sequence 2188, Ap  
96 154 4.0 2992 1 PCT-US03-26780-2188 Sequence 2257, Ap  
97 154 4.0 3065 1 PCT-US03-26780-2257 Sequence 240, App  
98 154 4.0 4873 28 US-10-236-177-240 Sequence 241, App  
99 154 4.0 4919 37 US-60-568-845-240 Sequence 240, App  
100 154 4.0 5314 37 US-60-568-845-240 Sequence 240, App

ALIGNMENTS

RESULT 1  
US-10-650-482-2  
; Sequence 2, Application US/10650482  
; GENERAL INFORMATION:  
; APPLICANT: Ron, David  
; TITLE OF INVENTION: METHODS OF SCREENING TEST COMPOUNDS USING GADD34L, AN eIF2alpha-S  
; TITLE OF INVENTION: PHOSPHATASE SUBUNIT  
; FILE REFERENCE: 5986/11712-US1  
; CURRENT APPLICATION NUMBER: US/10/650,482  
; CURRENT FILING DATE: 2003-08-28  
; PRIOR FILING DATE: 2003-08-28  
; PRIOR FILING DATE: 2003-08-28  
; NUMBER OF SEQ ID NOS: 10  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 2  
; LENGTH: 713  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-650-482-2  
Query Match 100.0%; Score 3804; DB 32; Length 713;  
Best Local Similarity 100.0%; Pred. No. 2.5e-237;  
Matches 713; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 MEPGTGSRKRLGPRAGFRFWPPPPRRSQAGSSKFPPTPLGPENSGNPTLLSSAQPETRV 60  
Db 1 MEPGTGSRKRLGPRAGFRFWPPPPRRSQAGSSKFPPTPLGPENSGNPTLLSSAQPETRV 60  
Qy 61 SYWTKLLSQLAPLPGLLQKVLINSQLFGGMFPTRWLDFAGVYSALRKALKREKPAAPTA 120  
Db 61 SYWTKLLSQLAPLPGLLQKVLINSQLFGGMFPTRWLDFAGVYSALRKALKREKPAAPTA 120  
Qy 121 QKSLSSQLDSSDPSVTSPLDWLEEGIHWOYSPDPLKLELKAKGSALDPAQAFLLEQOL 180  
Db 121 QKSLSSQLDSSDPSVTSPLDWLEEGIHWOYSPDPLKLELKAKGSALDPAQAFLLEQOL 180  
Qy 121 QKSLSSQLDSSDPSVTSPLDWLEEGIHWOYSPDPLKLELKAKGSALDPAQAFLLEQOL 180  
Db 121 QKSLSSQLDSSDPSVTSPLDWLEEGIHWOYSPDPLKLELKAKGSALDPAQAFLLEQOL 180

Qy 181 WGVELLSSQLSRLYSNRELSSPSGPNLQRIIDNFVSVYLLNPSYLDGPPRLVSYQN 240  
Db 181 WGVELLSSQLSRLYSNRELSSPSGPNLQRIIDNFVSVYLLNPSYLDGPPRLVSYQN 240  
Qy 241 SDGNSEVVGFTLTPESSCLREDHCHPOPLSAELIPASWGQCPPLSTEGPLPEIHLRMKR 300  
Db 241 SDGNSEVVGFTLTPESSCLREDHCHPOPLSAELIPASWGQCPPLSTEGPLPEIHLRMKR 300  
Qy 301 LEFLQANKGQDLPTPDQNGYHSLREHSLRMDPKHCRDNPQTFVPAAGDIPGNTQES 360  
Db 301 LEFLQANKGQDLPTPDQNGYHSLREHSLRMDPKHCRDNPQTFVPAAGDIPGNTQES 360  
Qy 361 TEEKIELLTTEVPALREESPESECPSEIPMEKEPGEGRISVVVDYSLYEGDLPI SARPA 420  
Db 361 TEEKIELLTTEVPALREESPESECPSEIPMEKEPGEGRISVVVDYSLYEGDLPI SARPA 420  
Qy 421 CSNKLIDYILGGASSDLETSDPEGEDWDEAEDDGFDSLSLSDLEODPEGLHLWNS 480  
Db 421 CSNKLIDYILGGASSDLETSDPEGEDWDEAEDDGFDSLSLSDLEODPEGLHLWNS 480  
Qy 481 FCSVDPNPNQFTATIQTAAIRIVPEEPSDSEKDLGKSDLENSQSGLPETPESHSGEE 540  
Db 481 FCSVDPNPNQFTATIQTAAIRIVPEEPSDSEKDLGKSDLENSQSGLPETPESHSGEE 540  
Qy 541 DMWESSADEAESLKLWNSFCNSDDPNPLNFKAPFQTSGENEKGRDSTKTPSEISVAISE 600  
Db 541 DMWESSADEAESLKLWNSFCNSDDPNPLNFKAPFQTSGENEKGRDSTKTPSEISVAISE 600  
Qy 601 CHTLLSCVKQLLSQESQECPSVQORVLSGGRHVVKKVTFLEEVTEYVYISGDRKRG 660  
Db 601 CHTLLSCVKQLLSQESQECPSVQORVLSGGRHVVKKVTFLEEVTEYVYISGDRKRG 660  
Qy 661 PWEEFARDGCRFKRIQETEDAIGYCLTFEHRERMFNRLOGTCFKGLNLVKQC 713  
Db 661 PWEEFARDGCRFKRIQETEDAIGYCLTFEHRERMFNRLOGTCFKGLNLVKQC 713  
RESULT 2  
US-10-170-205E-23916  
; Sequence 23916, Application US/10170205E  
; GENERAL INFORMATION:  
; APPLICANT: ADAMS, Mark  
; TITLE OF INVENTION: DEVICES, SUCH AS ARRAYS, COMPRISED OF HUMAN PROTEINS OR PROTEIN  
; TITLE OF INVENTION: CAPTURE AGENTS, AND USES THEREOF  
; FILE REFERENCE: CL001381  
; CURRENT APPLICATION NUMBER: US/10/170,205E  
; CURRENT FILING DATE: 2002-06-13  
; NUMBER OF SEQ ID NOS: 40312  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 23916  
; LENGTH: 713  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-170-205E-23916  
Query Match 99.9%; Score 3799; DB 27; Length 713;  
Best Local Similarity 99.9%; Pred. No. 6.4e-297;  
Matches 712; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 MEPGTGSRKRLGPRAGFRFWPPPPRRSQAGSSKFPPTPLGPENSGNPTLLSSAQPETRV 60  
Db 1 MEPGTGSRKRLGPRAGFRFWPPPPRRSQAGSSKFPPTPLGPENSGNPTLLSSAQPETRV 60  
Qy 61 SYWTKLLSQLAPLPGLLQKVLINSQLFGGMFPTRWLDFAGVYSALRKALKREKPAAPTA 120  
Db 61 SYWTKLLSQLAPLPGLLQKVLINSQLFGGMFPTRWLDFAGVYSALRKALKREKPAAPTA 120  
Qy 121 QKSLSSQLDSSDPSVTSPLDWLEEGIHWOYSPDPLKLELKAKGSALDPAQAFLLEQOL 180  
Db 121 QKSLSSQLDSSDPSVTSPLDWLEEGIHWOYSPDPLKLELKAKGSALDPAQAFLLEQOL 180  
Qy 181 WGVELLSSQLSRLYSNRELSSPSGPNLQRIIDNFVSVYLLNPSYLDGPPRLVSYQN 240  
Db 181 WGVELLSSQLSRLYSNRELSSPSGPNLQRIIDNFVSVYLLNPSYLDGPPRLVSYQN 240

```

Db 181  WGVLLPSSLSQRLYSNRELSSGPGPLNTQRIIDNFSVSVLLNPSYLDCCPRLEVSQN 240
Qy 241  SDGNSEVVGFTLTTPSSCLREDHCHPQLSAELIPASWQCPCPLSTEGPLPEIHILRMKR 300
Db 241  SDGNSEVVGFTLTTPSSCLREDHCHPQLSAELIPASWQCPCPLSTEGPLPEIHILRMKR 300
Qy 301  LEFLQANKGQDLPTPDQNGYHSLBEHSLLRMDPKHCRDNPQFVPAAGDIPGNTQES 360
Db 301  LEFLQANKGQDLPTPDQNGYHSLBEHSLLRMDPKHCRDNPQFVPAAGDIPGNTQES 360
Qy 361  TEEKIELLTTEVPLALEEESPSCPSSEIPMEKEPGEGRISVVDYSYLEGDLPI SARPA 420
Db 361  TEEKIELLTTEVPLALEEESPSCPSSEIPMEKEPGEGRISVVDYSYLEGDLPI SARPA 420
Qy 421  CSNKLIDYILGGASSDLETSDPEGEDWDEAEADGDFDSSLSLSDLEQDPGLHLWNS 480
Db 421  CSNKLIDYILGGASSDLETSDPEGEDWDEAEADGDFDSSLSLSDLEQDPGLHLWNS 480
Qy 481  FCSVDPNPNQFTATTIQTAAIRIVPEPSPDSKOLSGKSDLENSSQSGSLPETPHSGEE 540
Db 481  FCSVDPNPNQFTATTIQTAAIRIVPEPSPDSKOLSGKSDLENSSQSGSLPETPHSGEE 540
Qy 541  DWESSADEAESLKLWNSFCNSDDPNPLNFKAPFQTSGENEKGCRDSKTPESIVAISE 600
Db 541  DWESSADEAESLKLWNSFCNSDDPNPLNFKAPFQTSGENEKGCRDSKTPESIVAISE 600
Qy 601  CHTLLSKVQLLGSQSECPDSVQDVLSCGRHTRHVRKKTFLFEEVTEYIISGDEDRKG 660
Db 601  CHTLLSKVQLLGSQSECPDSVQDVLSCGRHTRHVRKKTFLFEEVTEYIISGDEDRKG 660
Qy 661  PWEEFARDGCRFKRIQETEDAIGYCLTTFEHRERMFNRLOGTCTPKGLNVLKQC 713
Db 661  PWEEFARDGCRFKRIQETEDAIGYCLTTFEHRERMFNRLOGTCTPKGLNVLKQC 713

RESULT 3
US-60-452-680-17655
; Sequence 17655, Application US/60452680
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; APPLICANT: GRUPE, Andrew
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE OF INVENTION: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01450
; CURRENT APPLICATION NUMBER: US/60-452,680
; CURRENT FILING DATE: 2003-03-07
; NUMBER OF SEQ ID NOS: 116213
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17655
; LENGTH: 713
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-452-680-17655

Query Match 99.9%; Score 3799; DB 37; Length 713;
Best Local Similarity 99.9%; Pred. No. 6.4e-297;
Matches 712; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1  MEPGTGSRKRLGRPRGFRFPWPPPPRRSQAGSKFPPTPLGPENSGNPTLLSSAQPETRV 60
Db 1  MEPGTGSRKRLGRPRGFRFPWPPPPRRSQAGSKFPPTPLGPENSGNPTLLSSAQPETRV 60
Qy 61  SYWTKLLSQLLAPLPGLLQKVLINSQFGMFPTRWLDFAGVYSALRALKRGKREPAAPTA 120
Db 61  SYWTKLLSQLLAPLPGLLQKVLINSQFGMFPTRWLDFAGVYSALRALKRGKREPAAPTA 120
Qy 121  QKSLSSQLDSDPSVTPDLWLEBEGHWHQVSPDPLKLELKAKGASLDPAQAFLLEQQL 180
Db 121  QKSLSSQLDSDPSVTPDLWLEBEGHWHQVSPDPLKLELKAKGASLDPAQAFLLEQQL 180
Qy 181  WGVLLPSSLSQRLYSNRELSSGPGPLNTQRIIDNFSVSVLLNPSYLDCCPRLEVSQN 240
Db 181  WGVLLPSSLSQRLYSNRELSSGPGPLNTQRIIDNFSVSVLLNPSYLDCCPRLEVSQN 240

```

```

Qy 241  SDGNSEVVGFTLTTPSSCLREDHCHPQLSAELIPASWQCPCPLSTEGPLPEIHILRMKR 300
Db 241  SDGNSEVVGFTLTTPSSCLREDHCHPQLSAELIPASWQCPCPLSTEGPLPEIHILRMKR 300
Qy 301  LEFLQANKGQDLPTPDQNGYHSLBEHSLLRMDPKHCRDNPQFVPAAGDIPGNTQES 360
Db 301  LEFLQANKGQDLPTPDQNGYHSLBEHSLLRMDPKHCRDNPQFVPAAGDIPGNTQES 360
Qy 361  TEEKIELLTTEVPLALEEESPSCPSSEIPMEKEPGEGRISVVDYSYLEGDLPI SARPA 420
Db 361  TEEKIELLTTEVPLALEEESPSCPSSEIPMEKEPGEGRISVVDYSYLEGDLPI SARPA 420
Qy 421  CSNKLIDYILGGASSDLETSDPEGEDWDEAEADGDFDSSLSLSDLEQDPGLHLWNS 480
Db 421  CSNKLIDYILGGASSDLETSDPEGEDWDEAEADGDFDSSLSLSDLEQDPGLHLWNS 480
Qy 481  FCSVDPNPNQFTATTIQTAAIRIVPEPSPDSKOLSGKSDLENSSQSGSLPETPHSGEE 540
Db 481  FCSVDPNPNQFTATTIQTAAIRIVPEPSPDSKOLSGKSDLENSSQSGSLPETPHSGEE 540
Qy 541  DWESSADEAESLKLWNSFCNSDDPNPLNFKAPFQTSGENEKGCRDSKTPESIVAISE 600
Db 541  DWESSADEAESLKLWNSFCNSDDPNPLNFKAPFQTSGENEKGCRDSKTPESIVAISE 600
Qy 601  CHTLLSKVQLLGSQSECPDSVQDVLSCGRHTRHVRKKTFLFEEVTEYIISGDEDRKG 660
Db 601  CHTLLSKVQLLGSQSECPDSVQDVLSCGRHTRHVRKKTFLFEEVTEYIISGDEDRKG 660
Qy 661  PWEEFARDGCRFKRIQETEDAIGYCLTTFEHRERMFNRLOGTCTPKGLNVLKQC 713
Db 661  PWEEFARDGCRFKRIQETEDAIGYCLTTFEHRERMFNRLOGTCTPKGLNVLKQC 713

RESULT 4
US-09-629-469A-18965
; Sequence 18965, Application US/09629469A
; GENERAL INFORMATION:
; APPLICANT: OTA, TOSHIO
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: NISHIKAWA, TETSUO
; APPLICANT: HAYASHI, KOJI
; APPLICANT: SAITO, KAOBU
; APPLICANT: YAMAMOTO, JUNICHI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: WAKAMATSU, AI
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: OTSUKI, TETSUJI
; TITLE OF INVENTION: PRIMERS FOR SYNTHESIZING FULL-LENGTH CDNA AND THEIR USE
; FILE REFERENCE: 084335/0123
; CURRENT APPLICATION NUMBER: US/09/629,469A
; CURRENT FILING DATE: 2000-07-28
; PRIOR APPLICATION NUMBER: JP 1999-248036
; PRIOR FILING DATE: 1999-07-29
; PRIOR APPLICATION NUMBER: JP 1999-300253
; PRIOR FILING DATE: 1999-08-27
; PRIOR APPLICATION NUMBER: JP 2000-118776
; PRIOR FILING DATE: 2000-01-11
; PRIOR APPLICATION NUMBER: JP 2000-183767
; PRIOR FILING DATE: 2000-05-02
; PRIOR APPLICATION NUMBER: JP 2000-241899
; PRIOR FILING DATE: 2000-06-09
; PRIOR APPLICATION NUMBER: 60/159,590
; PRIOR FILING DATE: 1999-10-18
; PRIOR APPLICATION NUMBER: 60/183,322
; PRIOR FILING DATE: 2000-02-17
; NUMBER OF SEQ ID NOS: 19025
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 18965
; LENGTH: 713
; TYPE: PRT
; ORGANISM: Homo sapiens

```

US-09-629-469A-18965

```

Query Match          99.7%; Score 3794; DB 20; Length 713;
Best Local Similarity 99.7%; Pred. No. 1.6e-296;
Matches 711; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MEGTGGSRKRLGPRAGFRFWPPFPFRSQAGSKFPTPLGPENSGNPTLLSSAQPETRV 60
Db 1 MEGTGGSRKRLGPRAGFRFWPPFPFRSQAGSKFPTPLGPENSGNPTLLSSAQPETRV 60

Qy 61 SYWTKLSQLLAPLGLQKVLWLSQLFQGGMPTRWLDVAGVYSALRKAGREKPAAPTA 120
Db 61 SYWTKLSQLLAPLGLQKVLWLSQLFQGGMPTRWLDVAGVYSALRKAGREKPAAPTA 120

Qy 121 QKSLSSQLDSSDPVSTPLDQNGYHSLWEEHSLRMDPKHCRDNPOTQFVPAAGDIPGNTQES 360
Db 121 QKSLSSQLDSSDPVSTPLDQNGYHSLWEEHSLRMDPKHCRDNPOTQFVPAAGDIPGNTQES 360

Qy 181 MGVELLPSSLOSRLYSNRELSSPGPLNIQRIDNFSVSVLLNPSYLDLCPRLLEVSYON 240
Db 181 MGVELLPSSLOSRLYSNRELSSPGPLNIQRIDNFSVSVLLNPSYLDLCPRLLEVSYON 240

Qy 241 SDGNSEVVGFTLTTPESCLREDHCHPQLSABLIPASWQCGLPSTEGLPETIHLRMKR 300
Db 241 SDGNSEVVGFTLTTPESCLREDHCHPQLSABLIPASWQCGLPSTEGLPETIHLRMKR 300

Qy 301 LEFLQKANKGQDLPTPDQNGYHSLWEEHSLRMDPKHCRDNPOTQFVPAAGDIPGNTQES 360
Db 301 LEFLQKANKGQDLPTPDQNGYHSLWEEHSLRMDPKHCRDNPOTQFVPAAGDIPGNTQES 360

Qy 361 TEEKIELLTTEVPLALEEESPEGCPSSEIPMEKEPGEGRISVVDYSYLEGDLPI SARPA 420
Db 361 TEEKIELLTTEVPLALEEESPEGCPSSEIPMEKEPGEGRISVVDYSYLEGDLPI SARPA 420

Qy 421 CSNKLIDYILGGASSDLTSSDPEGDWDEAEADGDFSDSLSDSDLEQDPEGLHLWNS 480
Db 421 CSNKLIDYILGGASSDLTSSDPEGDWDEAEADGDFSDSLSDSDLEQDPEGLHLWNS 480

Qy 481 FCSVDVPYNPQNFATTIQTAAIRIVPEEPSDEKDLGKSDLENSSQSGSLPETPHSSGEE 540
Db 481 FCSVDVPYNPQNFATTIQTAAIRIVPEEPSDEKDLGKSDLENSSQSGSLPETPHSSGEE 540

Qy 541 DWESSADEAESLKLWNSFCNSDDPNPLNFKAPFQTSGENEKGCRDCKTSPESIVAISE 600
Db 541 DWESSADEAESLKLWNSFCNSDDPNPLNFKAPFQTSGENEKGCRDCKTSPESIVAISE 600

Qy 601 CHTLLSCVKVLLGSQSECPDSVQRDVLGSGRTHVKKVTFLEEVTEYYISGDEDRKG 660
Db 601 CHTLLSCVKVLLGSQSECPDSVQRDVLGSGRTHVKKVTFLEEVTEYYISGDEDRKG 660

Qy 661 PWEFPARDGCRFQKRIQETEDAIGYCLTFEHRERMFNRLOQTGCFKGLNVLKQC 713
Db 661 PWEFPARDGCRFQKRIQETEDAIGYCLTFEHRERMFNRLOQTGCFKGLNVLKQC 713

```

RESULT 5

```

US-10-917-503-18965
; Sequence 18965, Application US/10917503
; GENERAL INFORMATION:
; APPLICANT: OTA, TOSHIO
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: NISHIKAWA, TETSUO
; APPLICANT: HAYASHI, KOJI
; APPLICANT: SAITO, KAORU
; APPLICANT: YAMAMOTO, JUNICHI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: WAKAMATSU, AI
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: OTSUKI, TETSUJI
; TITLE OF INVENTION: PRIMERS FOR SYNTHESIZING FULL-LENGTH CDNA AND THEIR USE
; FILE REFERENCE: 084335/0123
; CURRENT APPLICATION NUMBER: US/10/917,503

```

```

; CURRENT FILING DATE: 2004-08-13
; PRIOR APPLICATION NUMBER: US/09/629,469
; PRIOR FILING DATE: 2000-07-28
; PRIOR APPLICATION NUMBER: JP 1999-248036
; PRIOR FILING DATE: 1999-07-29
; PRIOR APPLICATION NUMBER: JP 1999-300253
; PRIOR FILING DATE: 1999-08-27
; PRIOR APPLICATION NUMBER: JP 2000-118776
; PRIOR FILING DATE: 2000-01-11
; PRIOR APPLICATION NUMBER: JP 2000-183767
; PRIOR FILING DATE: 2000-05-02
; PRIOR APPLICATION NUMBER: JP 2000-241899
; PRIOR FILING DATE: 2000-06-09
; PRIOR APPLICATION NUMBER: 60/159,590
; PRIOR FILING DATE: 1999-10-18
; PRIOR APPLICATION NUMBER: 60/183,322
; PRIOR FILING DATE: 2000-02-17
; NUMBER OF SEQ ID NOS: 19025
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 18965
; LENGTH: 713
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-917-503-18965

Query Match          99.7%; Score 3794; DB 35; Length 713;
Best Local Similarity 99.7%; Pred. No. 1.6e-296;
Matches 711; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MEGTGGSRKRLGPRAGFRFWPPFPFRSQAGSKFPTPLGPENSGNPTLLSSAQPETRV 60
Db 1 MEGTGGSRKRLGPRAGFRFWPPFPFRSQAGSKFPTPLGPENSGNPTLLSSAQPETRV 60

Qy 61 SYWTKLSQLLAPLGLQKVLWLSQLFQGGMPTRWLDVAGVYSALRKAGREKPAAPTA 120
Db 61 SYWTKLSQLLAPLGLQKVLWLSQLFQGGMPTRWLDVAGVYSALRKAGREKPAAPTA 120

Qy 121 QKSLSSQLDSSDPVSTPLDQNGYHSLWEEHSLRMDPKHCRDNPOTQFVPAAGDIPGNTQES 360
Db 121 QKSLSSQLDSSDPVSTPLDQNGYHSLWEEHSLRMDPKHCRDNPOTQFVPAAGDIPGNTQES 360

Qy 181 MGVELLPSSLOSRLYSNRELSSPGPLNIQRIDNFSVSVLLNPSYLDLCPRLLEVSYON 240
Db 181 MGVELLPSSLOSRLYSNRELSSPGPLNIQRIDNFSVSVLLNPSYLDLCPRLLEVSYON 240

Qy 241 SDGNSEVVGFTLTTPESCLREDHCHPQLSABLIPASWQCGLPSTEGLPETIHLRMKR 300
Db 241 SDGNSEVVGFTLTTPESCLREDHCHPQLSABLIPASWQCGLPSTEGLPETIHLRMKR 300

Qy 301 LEFLQKANKGQDLPTPDQNGYHSLWEEHSLRMDPKHCRDNPOTQFVPAAGDIPGNTQES 360
Db 301 LEFLQKANKGQDLPTPDQNGYHSLWEEHSLRMDPKHCRDNPOTQFVPAAGDIPGNTQES 360

Qy 361 TEEKIELLTTEVPLALEEESPEGCPSSEIPMEKEPGEGRISVVDYSYLEGDLPI SARPA 420
Db 361 TEEKIELLTTEVPLALEEESPEGCPSSEIPMEKEPGEGRISVVDYSYLEGDLPI SARPA 420

Qy 421 CSNKLIDYILGGASSDLTSSDPEGDWDEAEADGDFSDSLSDSDLEQDPEGLHLWNS 480
Db 421 CSNKLIDYILGGASSDLTSSDPEGDWDEAEADGDFSDSLSDSDLEQDPEGLHLWNS 480

Qy 481 FCSVDVPYNPQNFATTIQTAAIRIVPEEPSDEKDLGKSDLENSSQSGSLPETPHSSGEE 540
Db 481 FCSVDVPYNPQNFATTIQTAAIRIVPEEPSDEKDLGKSDLENSSQSGSLPETPHSSGEE 540

Qy 541 DWESSADEAESLKLWNSFCNSDDPNPLNFKAPFQTSGENEKGCRDCKTSPESIVAISE 600
Db 541 DWESSADEAESLKLWNSFCNSDDPNPLNFKAPFQTSGENEKGCRDCKTSPESIVAISE 600

Qy 601 CHTLLSCVKVLLGSQSECPDSVQRDVLGSGRTHVKKVTFLEEVTEYYISGDEDRKG 660
Db 601 CHTLLSCVKVLLGSQSECPDSVQRDVLGSGRTHVKKVTFLEEVTEYYISGDEDRKG 660

```

Qy 661 PWEEFARDGCRFQKRIQETEDAIGYCLTFFHRRMFNRLOGTCTFKGLNVLKQC 713  
Db 661 PWEEFARDGCRFQKRIQETEDAIGYCLTFFHRRMFNRLOGTCTFKGLNVLKQC 713

## RESULT 6

US-09-757-028-2348  
; Sequence 2348, Application US/09757028  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PM001  
; CURRENT APPLICATION NUMBER: US/09/757,028  
; CURRENT FILING DATE: 2001-01-09  
; PRIOR APPLICATION NUMBER: 60/179,065  
; PRIOR FILING DATE: 2000-01-31  
; PRIOR APPLICATION NUMBER: 60/180,628  
; PRIOR FILING DATE: 2000-02-04  
; NUMBER OF SEQ ID NOS: 2660  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2348  
; LENGTH: 720  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-757-028-2348

Query Match 99.6%; Score 3789; DB 22; Length 720;  
Best Local Similarity 99.7%; Pred. No. 4.2e-296;  
Matches 711; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MEPTGGSRKRLGPRAGFRFWPPFFRRRSQAGSSKFTPLGPENSGNPTLLSSAQPETRV 60  
Db 8 MEPTGGSRKRLGPRAGFRFWPPFFRRRSQAGSSKFTPLGPENSGNPTLLSSAQPETRV 67  
Qy 61 SYWTKLSQLLAPLPGLLQKVLWSQLFGGMFTRWLDFAGVYSALRALKGRKPAAPTA 120  
Db 68 SYWTKLSQLLAPLPGLLQKVLWSQLFGGMFTRWLDFAGVYSALRALKGRKPAAPTA 127  
Qy 121 QKSLSLQDSSDPSVTSPLDWEEGHWOYSPDCLKELKAKGSALDPAQAFLLEQQ 180  
Db 128 QKSLSLQDSSDPSVTSPLDWEEGHWOYSPDCLKELKAKGSALDPAQAFLLEQQ 187  
Qy 181 WGVLELLPSSLOSRLYSNRELSSPSGLNIQRIIDNFSVSVLLNPSYLDLCPPLREVSQN 240  
Db 188 WGVLELLPSSLOSRLYSNRELSSPSGLNIQRIIDNFSVSVLLNPSYLDLCPPLREVSQN 247  
Qy 241 SDGNSEVVGFTLTTPESSCLREDHCHPQLSABLIPASWQCPLSTEGLPETIHLRMKR 300  
Db 248 SDGNSEVVGFTLTTPESSCLREDHCHPQLSABLIPASWQCPLSTEGLPETIHLRMKR 307  
Qy 301 LEFLQANKGQDLPTPDQNGYHSLBEEHSLLRMDPKHCRDNPTQFVPAAGDIPGNTQES 360  
Db 308 LEFLQANKGQDLPTPDQNGYHSLBEEHSLLRMDPKHCRDNPTQFVPAAGDIPGNTQES 367  
Qy 361 TEKIELLTTEVPALAEESPSGCPSEIPMEKEPGEGRISVVDYSYLEGDLPI SARPA 420  
Db 368 TEKIELLTTEVPALAEESPSGCPSEIPMEKEPGEGRISVVDYSYLEGDLPI SARPA 427  
Qy 421 CSNKLIDYILGGASSDLETSSDPGEWDDEAEDGFDSDSLSDLEQDPGHLWNS 480  
Db 428 CSNKLIDYILGGASSDLETSSDPGEWDDEAEDGFDSDSLSDLEQDPGHLWNS 487  
Qy 481 FCSVDYPNPQNTATTQTAARIIVPEEPSDSKOLSGKSDLENSQSGSLPETPHSSGEE 540  
Db 488 FCSVDYPNPQNTATTQTAARIIVPEEPSDSKOLSGKSDLENSQSGSLPETPHSSGEE 547  
Qy 541 DDWESSADEAESLKLWNSFCNSDDPYNPLNFKAPFQTSGENEKCRCDSKTPSPESIVASE 600  
Db 548 DDWESSADEAESLKLWNSFCNSDDPYNPLNFKAPFQTSGENEKCRCDSKTPSPESIVASE 607  
Qy 601 CHTLLSCKVQLLGQSEECPSVQDVLSGGRHHTVKKKVTFLFEEVTEYIISGDEDRKG 660  
Db 608 CHTLLSCKVQLLGQSEECPSVQDVLSGGRHHTVKKKVTFLFEEVTEYIISGDEDRKG 667

Qy 661 PWEEFARDGCRFQKRIQETEDAIGYCLTFFHRRMFNRLOGTCTFKGLNVLKQC 713  
Db 668 PWEEFARDGCRFQKRIQETEDAIGYCLTFFHRRMFNRLOGTCTFKGLNVLKQC 720

## RESULT 7

US-10-222-911-2348  
; Sequence 2348, Application US/10222911  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PM001CIN  
; CURRENT APPLICATION NUMBER: US/10/222,911  
; CURRENT FILING DATE: 2002-08-19  
; PRIOR APPLICATION NUMBER: 09/757,028  
; PRIOR FILING DATE: 2001-01-09  
; PRIOR APPLICATION NUMBER: 60/179,065  
; PRIOR FILING DATE: 2000-01-31  
; PRIOR APPLICATION NUMBER: 60/180,628  
; PRIOR FILING DATE: 2000-02-04  
; NUMBER OF SEQ ID NOS: 2660  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2348  
; LENGTH: 720  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-222-911-2348

Query Match 99.6%; Score 3789; DB 28; Length 720;  
Best Local Similarity 99.7%; Pred. No. 4.2e-296;  
Matches 711; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MEPTGGSRKRLGPRAGFRFWPPFFRRRSQAGSSKFTPLGPENSGNPTLLSSAQPETRV 60  
Db 8 MEPTGGSRKRLGPRAGFRFWPPFFRRRSQAGSSKFTPLGPENSGNPTLLSSAQPETRV 67  
Qy 61 SYWTKLSQLLAPLPGLLQKVLWSQLFGGMFTRWLDFAGVYSALRALKGRKPAAPTA 120  
Db 68 SYWTKLSQLLAPLPGLLQKVLWSQLFGGMFTRWLDFAGVYSALRALKGRKPAAPTA 127  
Qy 121 QKSLSLQDSSDPSVTSPLDWEEGHWOYSPDCLKELKAKGSALDPAQAFLLEQQ 180  
Db 128 QKSLSLQDSSDPSVTSPLDWEEGHWOYSPDCLKELKAKGSALDPAQAFLLEQQ 187  
Qy 181 WGVLELLPSSLOSRLYSNRELSSPSGLNIQRIIDNFSVSVLLNPSYLDLCPPLREVSQN 240  
Db 188 WGVLELLPSSLOSRLYSNRELSSPSGLNIQRIIDNFSVSVLLNPSYLDLCPPLREVSQN 247  
Qy 241 SDGNSEVVGFTLTTPESSCLREDHCHPQLSABLIPASWQCPLSTEGLPETIHLRMKR 300  
Db 248 SDGNSEVVGFTLTTPESSCLREDHCHPQLSABLIPASWQCPLSTEGLPETIHLRMKR 307  
Qy 301 LEFLQANKGQDLPTPDQNGYHSLBEEHSLLRMDPKHCRDNPTQFVPAAGDIPGNTQES 360  
Db 308 LEFLQANKGQDLPTPDQNGYHSLBEEHSLLRMDPKHCRDNPTQFVPAAGDIPGNTQES 367  
Qy 361 TEKIELLTTEVPALAEESPSGCPSEIPMEKEPGEGRISVVDYSYLEGDLPI SARPA 420  
Db 368 TEKIELLTTEVPALAEESPSGCPSEIPMEKEPGEGRISVVDYSYLEGDLPI SARPA 427  
Qy 421 CSNKLIDYILGGASSDLETSSDPGEWDDEAEDGFDSDSLSDLEQDPGHLWNS 480  
Db 428 CSNKLIDYILGGASSDLETSSDPGEWDDEAEDGFDSDSLSDLEQDPGHLWNS 487  
Qy 481 FCSVDYPNPQNTATTQTAARIIVPEEPSDSKOLSGKSDLENSQSGSLPETPHSSGEE 540  
Db 488 FCSVDYPNPQNTATTQTAARIIVPEEPSDSKOLSGKSDLENSQSGSLPETPHSSGEE 547  
Qy 541 DDWESSADEAESLKLWNSFCNSDDPYNPLNFKAPFQTSGENEKCRCDSKTPSPESIVASE 600  
Db 548 DDWESSADEAESLKLWNSFCNSDDPYNPLNFKAPFQTSGENEKCRCDSKTPSPESIVASE 607

QY 601 CHTLLSKVQLGSGQSECPDSVQRDVLSSGRHHTVKKRVTFLEVTYYISGDEDRKG 660  
DB 608 CHTLLSKVQLGSGQSECPDSVQRDVLSSGRHHTVKKRVTFLEVTYYISGDEDRKG 667  
QY 661 PWEFARDGCRFQKRIQIETEDAIGYCLTFEHRMFRNLQTCFGLNLVKQC 713  
DB 668 PWEFARDGCRFQKRIQIETEDAIGYCLTFEHRMFRNLQTCFGLNLVKQC 720

RESULT 8  
PCT-US00-05883-941  
; Sequence 941, Application PC/TUS0005883  
; GENERAL INFORMATION:  
; APPLICANT: Craig Rosen,  
; APPLICANT: Steve Ruben,  
; TITLE OF INVENTION: Human Colon Cancer Associated Gene Sequences and Polypeptides  
; FILE REFERENCE: PA102PCT  
; CURRENT APPLICATION NUMBER: PCT/US00/05883  
; CURRENT FILING DATE: 2000-03-08  
; EARLIER APPLICATION NUMBER: 60/124,270  
; EARLIER FILING DATE: 1999-03-12  
; NUMBER OF SEQ ID NOS: 1556  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 941  
; LENGTH: 707  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (265)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (271)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (307)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
PCT-US00-05883-941

Query Match 95.3%; Score 3627; DB 1; Length 707;  
Best Local Similarity 98.8%; Pred. No. 5e-283;  
Matches 683; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 23 PFFPRRSQAGSKFPTPLGPNSEGNPTLLSSAQPETRVSYWTKLSQLLAPLPGLLQKVL 82  
DB 17 PSVSRRSQAGSKFPTPLGPNSEGNPTLLSSAQPETRVSYWTKLSQLLAPLPGLLQKVL 76  
QY 83 IWSQLFGCMFPTRWLDFAGVYSALRALKGRKPAAPTAQKSLSSQLDSSDPSVTSPLDW 142  
DB 77 IWSQLFGCMFPTRWLDFAGVYSALRALKGRKPAAPTAQKSLSSQLDSSDPSVTSPLDW 136  
QY 143 LEEGHWOYSPDDLKLELKAKGSALDPAQAFLLEQQLWGVELLPSLSQRLYSNREIGS 202  
DB 137 LEEGHWOYSPDDLKLELKAKGSALDPAQAFLLEQQLWGVELLPSLSQRLYSNREIGS 196  
QY 203 SPSPGLNIQRIDNFSVSYLLNPSYLDCCFPRLEVSQNSDGNSEVVGFTLTPESSCLRE 262  
DB 197 SPSPGLNIQRIDNFSVSYLLNPSYLDCCFPRLEVSQNSDGNSEVVGFTLTPESSCLRE 256  
QY 263 DHCHPQPLSAELIPASWQGCPLSTEGPLPEIHHLRMKLEFLQKASKQDXPTPDQNGY 322  
DB 257 DHCHPQPLSAELIPASWQGCPLSTEGPLPEIHHLRMKLEFLQKASKQDXPTPDQNGY 316  
QY 323 HSLEEHSLRLMDPKHCRDNTQFVPAAGDIPGNTQESTEEKIELLTTEVPALAEESPS 382  
DB 317 HSLEEHSLRLMDPKHCRDNTQFVPAAGDIPGNTQESTEEKIELLTTEVPALAEESPS 376  
QY 383 EGCSPSEIPMEKEPGEGRISVVDYSYLEGDLPI SARPACSNKLDYILGGASSDLETSSD 442  
DB 377 EGCSPSEIPMEKEPGEGRISVVDYSYLEGDLPI SARPACSNKLDYILGGASSDLETSSD 436

QY 443 PEGEDWDEAEADDGFDSDSSLSQSDLEQDPEGLHLMNSFCSDVPYNPQNFTATTIQTAAARI 502  
DB 437 PEGEDWDEAEADDGFDSDSSLSQSDLEQDPEGLHLMNSFCSDVPYNPQNFTATTIQTAAARI 496  
QY 503 VPEEPSSEKDLGKSKDLENSSQSGSLPETPEHSSGDEDDWESSADEASLKLNSEFCNS 562  
DB 497 VPEEPSSEKDLGKSKDLENSSQSGSLPETPEHSSGDEDDWESSADEASLKLNSEFCNS 556  
QY 563 DDYNPLNFKAPFOTSGENEGKCRDSTKTPSESVIAISECHTLLSKVQLGSGQSECPDS 622  
DB 557 DDYNPLNFKAPFOTSGENEGKCRDSTKTPSESVIAISECHTLLSKVQLGSGQSECPDS 616  
QY 623 VQRDVLSSGRHHTVKKRVTFLEVTYYISGDEDRKGPWEEFARDGCRFQKRIQIETEDA 682  
DB 617 VQRDVLSSGRHHTVKKRVTFLEVTYYISGDEDRKGPWEEFARDGCRFQKRIQIETEDA 676  
QY 683 IGYCLTFEHRMFRNLQTCFGLNLVKQC 713  
DB 677 IGYCLTFEHRMFRNLQTCFGLNLVKQC 707

RESULT 9  
US-09-925-299-941  
; Sequence 941, Application US/09925299  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
; FILE REFERENCE: PA102  
; CURRENT APPLICATION NUMBER: US/09/925,299  
; CURRENT FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: PCT/US00/05883  
; PRIOR FILING DATE: 2000-03-08  
; PRIOR APPLICATION NUMBER: 60/124,270  
; PRIOR FILING DATE: 1999-03-12  
; NUMBER OF SEQ ID NOS: 1556  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 941  
; LENGTH: 707  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (265)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (271)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (307)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-925-299-941

Query Match 95.3%; Score 3627; DB 24; Length 707;  
Best Local Similarity 98.8%; Pred. No. 5e-283;  
Matches 683; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 23 PFFPRRSQAGSKFPTPLGPNSEGNPTLLSSAQPETRVSYWTKLSQLLAPLPGLLQKVL 82  
DB 17 PSVSRRSQAGSKFPTPLGPNSEGNPTLLSSAQPETRVSYWTKLSQLLAPLPGLLQKVL 76  
QY 83 IWSQLFGCMFPTRWLDFAGVYSALRALKGRKPAAPTAQKSLSSQLDSSDPSVTSPLDW 142  
DB 77 IWSQLFGCMFPTRWLDFAGVYSALRALKGRKPAAPTAQKSLSSQLDSSDPSVTSPLDW 136  
QY 143 LEEGHWOYSPDDLKLELKAKGSALDPAQAFLLEQQLWGVELLPSLSQRLYSNREIGS 202  
DB 137 LEEGHWOYSPDDLKLELKAKGSALDPAQAFLLEQQLWGVELLPSLSQRLYSNREIGS 196  
QY 203 SPSPGLNIQRIDNFSVSYLLNPSYLDCCFPRLEVSQNSDGNSEVVGFTLTPESSCLRE 262  
DB 197 SPSPGLNIQRIDNFSVSYLLNPSYLDCCFPRLEVSQNSDGNSEVVGFTLTPESSCLRE 256  
QY 263 DHCHPQPLSAELIPASWQGCPLSTEGPLPEIHHLRMKLEFLQKASKQDXPTPDQNGY 322

Db 257 DHCHPOPLXAEIPXSGCPPLSTGLPBIHLRMKRLBFLQOASGXQDPTPDQNGY 316  
Qy 323 HSLREHSLRMDPKHCRDNPQVPAAGDI PGNTQESTBEKTELLTTEVPLALEBSPS 382  
Db 317 HSLREHSLRMDPKHCRDNPQVPAAGDI PGNTQESTBEKTELLTTEVPLALEBSPS 376  
Qy 383 EGCPSSEIPMEKEFEGEGRISVVDYSYLEGDLPI SARPAKSNKLDIYILGASSDLETSSD 442  
Db 377 EGCPSSEIPMEKEFEGEGRISVVDYSYLEGDLPI SARPAKSNKLDIYILGASSDLETSSD 436  
Qy 443 PEGEDWDEABDDGFDSDSLSDLEQDPGEGHLWNSFCSDVPYNPONTTATIQTAARI 502  
Db 437 PEGEDWDEABDDGFDSDSLSDLEQDPGEGHLWNSFCSDVPYNPONTTATIQTAARI 496  
Qy 503 VPEEPSDEKDLGSKSDLENSGSLPETPEHSSGDEDDWESSADEASLKLWNSFCNS 562  
Db 497 VPEEPSDEKDLGSKSDLENSGSLPETPEHSSGDEDDWESSADEASLKLWNSFCNS 556  
Qy 563 DDYPNPLNFKAPFOTSGENEGKCRDSTPSESIVAI SECHTLILSKVQLLGSQSECPDS 622  
Db 557 DDYPNPLNFKAPFOTSGENEGKCRDSTPSESIVAI SECHTLILSKVQLLGSQSECPDS 616  
Qy 623 VORDVLSGGRHTRVKKVTFLEEVTEYIISGDEDRKGPWEEFARDCRCRQKRIQETEDA 682  
Db 617 VORDVLSGGRHTRVKKVTFLEEVTEYIISGDEDRKGPWEEFARDCRCRQKRIQETEDA 676  
Qy 683 IGYCLTFEHRERMFNRLOQTCFKGLNLVKQC 713  
Db 677 IGYCLTFEHRERMFNRLOQTCFKGLNLVKQC 707

RESULT 10  
US-10-650-482-4  
; Sequence 4, Application US/10650482  
; GENERAL INFORMATION:  
; APPLICANT: Ron, David  
; TITLE OF INVENTION: METHODS OF SCREENING TEST COMPOUNDS USING GADD34L, AN eIF2alpha-S  
; FILE REFERENCE: 5986/11712-US1  
; CURRENT APPLICATION NUMBER: US/10/650,482  
; PRIOR FILING DATE: 2003-08-28  
; PRIOR APPLICATION NUMBER: US 60/408,679  
; NUMBER OF SEQ ID NOS: 10  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 4  
; LENGTH: 698  
; TYPE: PRT  
; ORGANISM: Mus musculus  
US-10-650-482-4

Query Match 58.4%; Score 2223; DB 32; Length 698;  
Best Local Similarity 65.0%; Pred. No. 1.4e-169;  
Matches 456; Conservative 53; Mismatches 174; Indels 18; Gaps 7;  
Qy 1 MEPTGSGRKLGRAGFRWPPFPFRSQAGSSKPTPLGPNSGNPTLLSSAQPTRV 60  
Db 1 METGTHARKRPGRLGSRWRLPFL-RRSHACSEFPFPPSRQPNGN-----SALPERRT 54  
Qy 61 SYWTKLSQLLAPLPGLLQKVLVWSQLFGGMFPTRWLDFAVYSALRALKGRSKPAAPTA 120  
Db 55 RYWTKLJSQLLALPSLFQKLLWSQLSGGLIPTRWLDFAASYSALRASRGRESDAPTV 114  
Qy 121 QKSLSSQLD-SSDPSTVSPDLWLEBGIHWQYSPDPLKELKAKGSALDPAQAFLLEQQ 179  
Db 115 QKSLSYTAAGLFAKTRVVTALARGGTPAVLVLRLEVKLKAQERALDSPAAPTFLEEQ 174  
Qy 180 LWGVLLPSSLOSRLYNRELSSGSPGPIAQRIDNFSVSYLLNPSYLDGCFRLEVSQ 239  
Db 175 LWGVLLPSSLQAGLVSHRELSSGSPGPIAQRIDNFSVSYLLNPSYLDYLPQLGURCQ 234

Qy 240 NSDGNSEVVGFOPLTPESSCLREDHCHPOPLXAEIPASWQGCCPPLSTGLPBIHLRMK 299  
Db 235 SSAGGQFVGFRTLTTPESCYLSDGCHPOPLXAEISATARRCPPLSTGLPBIHRRMR 294  
Qy 300 RLEFLQOANKGODLPTPDONGVHSLREHSLRMDPKHCRDNPQVPAAGDI PGNTQ 359  
Db 295 WLVEL-OPNQGDLPFLDQNGVHSLREHSLRMDPKHCRDNPQVPAAGDI PGNTQ 349  
Qy 360 STEEKTELLTTEVPLALEBSPSEIPMEKEFEGEGRISVVDYSYLEGDLPI SARPA 419  
Db 350 PTEKKPELVQEV-----SQSPQSSLPFCLEPVEKECEEDHTNATDLSDRGESLPVSTRP 404  
Qy 420 ACSNKLDIYILGASSDLETSDPEGEDWDEABDDGFDSDSLSDLEQDPGEGHLWNS 479  
Db 405 VCSNKLDIYILGASSDLETSDPEGEDWDEABDDGFDSDSLSDLEQDPGEGHLWNS 464  
Qy 480 SFCSDVPYNPONTTATIQTAARI VPEEPSDEKDLGSKSDLENSGSLPETPEHSSG 539  
Db 465 SFHSDVPYNPONTTATIQTAARI PRDPSDSTGTSWSGCGV-GSCQEGPLPETPDHSSG 523  
Qy 540 EDDWESSADEASLKLWNSFCSDVPYNPLNFKAPFOTSGENEGKCRDSTPSESIVAI 599  
Db 524 EDDWESSADEASLKLWNSFCSDVPYNPLNFKAPFOTSGENEGKCRDSTPSESIVAI 583  
Qy 600 ECHTLILSKVQLLGSQSECPDSVQRDVLSGGRHTRVKKVTFLEEVTEYIISGDEDRK 659  
Db 584 GHHTLILSKVQLLGSQSECPDSVQRDVLSGGRHTRVKKVTFLEEVTEYIISGDEDRK 643  
Qy 660 GPWEEFARDCRCRQKRIQETEDAIGYCLTFEHRERMFNRLO 700  
Db 644 GPWEEFARDCRCRQKRIQETEDAIGYCLTFEHRERMFNRLO 684

RESULT 11  
PCT-US00-35017A-963  
; Sequence 963, Application PC/TUS0035017A  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq Inc  
; TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides  
; FILE REFERENCE: 784PCT  
; CURRENT APPLICATION NUMBER: PCT/US00/35017A  
; CURRENT FILING DATE: 2000-12-22  
; PRIOR FILING DATE: 2000-01-21  
; PRIOR APPLICATION NUMBER: US09/488,725  
; PRIOR FILING DATE: 2000-01-21  
; PRIOR APPLICATION NUMBER: US09/552,317  
; NUMBER OF SEQ ID NOS: 1478  
; SEQ ID NO 963  
; LENGTH: 153  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
PCT-US00-35017A-963

Query Match 21.4%; Score 813; DB 1; Length 153;  
Best Local Similarity 100.0%; Pred. No. 9.9e-57;  
Matches 150; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 564 DPYNPLNFKAPFOTSGENEGKCRDSTPSESIVAI SECHTLILSKVQLLGSQSECPDSV 623  
Db 4 DPYNPLNFKAPFOTSGENEGKCRDSTPSESIVAI SECHTLILSKVQLLGSQSECPDSV 63  
Qy 624 QRDVLSGGRHTRVKKVTFLEEVTEYIISGDEDRKGPWEEFARDCRCRQKRIQETEDA 693  
Db 64 QRDVLSGGRHTRVKKVTFLEEVTEYIISGDEDRKGPWEEFARDCRCRQKRIQETEDA 123  
Qy 684 GYCLTFEHRERMFNRLOQTCFKGLNLVKQC 713  
Db 124 GYCLTFEHRERMFNRLOQTCFKGLNLVKQC 153

RESULT 12  
US-10-296-115-963  
; Sequence 963, Application US/10296115



```
; GENERAL INFORMATION:
; APPLICANT: Hyseq Inc
; TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides
; FILE REFERENCE: 784PCT
; CURRENT APPLICATION NUMBER: US/10/296,115
; CURRENT FILING DATE: 2002-11-18
; PRIOR APPLICATION NUMBER: US09/488,725
; PRIOR FILING DATE: 2000-01-21
; PRIOR APPLICATION NUMBER: US09/552,317
; PRIOR FILING DATE: 2000-04-25
; NUMBER OF SEQ ID NOS: 1478
; SEQ ID NO 963
; LENGTH: 153
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-296-115-963

Query Match      21.4%; Score 813; DB 28; Length 153;
Best Local Similarity 100.0%; Pred. No. 9.9e-57;
Matches 150; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      564 DPNPLNFKAPQTSGENEKGRDSTPESIVAISECHTLISCKVQLLGSQSECPDSV 623
Db      4 DPNPLNFKAPQTSGENEKGRDSTPESIVAISECHTLISCKVQLLGSQSECPDSV 63

Qy      624 QRDVLSCGRHTHVKKKVTFLVEEVTYYISGDEDRKGPWFEEFARDGCRFQKRIQETDAI 683
Db      64 QRDVLSCGRHTHVKKKVTFLVEEVTYYISGDEDRKGPWFEEFARDGCRFQKRIQETDAI 123

Qy      684 GYCLTFEHRMFMNRLQGTCTCFKGLNLVKQC 713
Db      124 GYCLTFEHRMFMNRLQGTCTCFKGLNLVKQC 153

RESULT 13
PCT-US01-14827-15743
; Sequence 15743, Application PC/TUS0114827
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 21272-104
; CURRENT APPLICATION NUMBER: PCT/US01/14827
; CURRENT FILING DATE: 2001-05-16
; PRIOR APPLICATION NUMBER: 09/577,408
; PRIOR FILING DATE: 2000-05-18
; NUMBER OF SEQ ID NOS: 16102
; SOFTWARE: Custom
; SEQ ID NO 15743
; LENGTH: 101
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US01-14827-15743

Query Match      13.4%; Score 511; DB 1; Length 101;
Best Local Similarity 95.0%; Pred. No. 1.3e-32;
Matches 95; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy      614 SQSECPDSVQDVLSCGRHTHVKKKVTFLVEEVTYYISGDEDRKGPWFEEFARDGCRFQ 673
Db      2 SQSECPDSVQDVLSCGRHTHVKKKVTFLVEEVTYYISGDEDRKGPWFEEFARDGCSVQ 61

Qy      674 KRIQETDAIGYCLTFEHRMFMNRLQGTCTCFKGLNLVKQC 713
Db      62 KRIQETDAIGYCLTFEHRMFMNRLQGTCTCFKGLNLVKQC 101

RESULT 14
US-60-177-571-4491
; Sequence 4491, Application US/60177571
; GENERAL INFORMATION:
; APPLICANT: Bonazzi, Vivien
; TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS,
```

```
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL000201
; CURRENT APPLICATION NUMBER: US/60/177,571
; CURRENT FILING DATE: 2000-01-27
; NUMBER OF SEQ ID NOS: 5082
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4491
; LENGTH: 73
; TYPE: PRT
; ORGANISM: HUMAN
US-60-177-571-4491

Query Match      10.6%; Score 402; DB 37; Length 73;
Best Local Similarity 100.0%; Pred. No. 4.8e-24;
Matches 73; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      641 VTFLVEEVTYYISGDEDRKGPWFEEFARDGCRFQKRIQETDAIGYCLTFEHRMFMNRLQ 700
Db      1 VTFLVEEVTYYISGDEDRKGPWFEEFARDGCRFQKRIQETDAIGYCLTFEHRMFMNRLQ 60

Qy      701 GTCFKGLNLVKQC 713
Db      61 GTCFKGLNLVKQC 73

RESULT 15
US-60-162-247-4518
; Sequence 4518, Application US/60162247
; GENERAL INFORMATION:
; APPLICANT: Bonazzi, Vivien
; TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: CL000127
; CURRENT APPLICATION NUMBER: US/60/162,247
; CURRENT FILING DATE: 1999-10-29
; NUMBER OF SEQ ID NOS: 5442
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4518
; LENGTH: 73
; TYPE: PRT
; ORGANISM: Human
US-60-162-247-4518

Query Match      10.5%; Score 399; DB 37; Length 73;
Best Local Similarity 98.6%; Pred. No. 8.4e-24;
Matches 72; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy      641 VTFLVEEVTYYISGDEDRKGPWFEEFARDGCRFQKRIQETDAIGYCLTFEHRMFMNRLQ 700
Db      1 VTFLVEEVTYYISGDEDRKGPWFEEFARDGCRFQKRIQETDAIGYCLTFDHRMFMNRLQ 60

Qy      701 GTCFKGLNLVKQC 713
Db      61 GTCFKGLNLVKQC 73

Search completed: September 16, 2005, 10:53:27
Job time : 189.419 secs
```



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 16, 2005, 10:42:43 ; Search time 86.4089 Seconds  
(without alignments)  
1348.829 Million cell updates/sec

Title: US-10-650-482-2

Perfect score: 3804

Sequence: 1 MEPTGCGSKRLGPRAGFRP.....RMFNRLLQGTCKGLNVLKQC 713

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 760079 seqs, 163465437 residues

Total number of hits satisfying chosen parameters: 760079

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

Pending Patents\_AA New:\*  
1: /cgn2\_6/ptodata/1/paa/PCT\_NEW\_COMB.pep.\*  
2: /cgn2\_6/ptodata/1/paa/US06\_NEW\_COMB.pep.\*  
3: /cgn2\_6/ptodata/1/paa/US07\_NEW\_COMB.pep.\*  
4: /cgn2\_6/ptodata/1/paa/US08\_NEW\_COMB.pep.\*  
5: /cgn2\_6/ptodata/1/paa/US09\_NEW\_COMB.pep.\*  
6: /cgn2\_6/ptodata/1/paa/US10\_NEW\_COMB.pep.\*  
7: /cgn2\_6/ptodata/1/paa/US11\_NEW\_COMB.pep.\*  
8: /cgn2\_6/ptodata/1/paa/US60\_NEW\_COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	260.5	6.8	674	7	US-11-040-219-2
2	195	5.1	605	7	US-11-040-219-5
3	167	4.4	1596	7	US-11-060-005-4
4	160.5	4.2	1562	6	US-10-450-763-57852
5	160.5	4.2	1702	6	US-10-450-763-38864
6	158.5	4.2	1585	6	US-10-990-328A-9765
7	158.5	4.2	5303	7	US-11-097-143-30390
8	158	4.2	2768	7	US-11-097-143-31983
9	157	4.1	2766	6	US-10-877-346-62
10	156.5	4.1	1503	1	PCT-US05-18850-797
11	156	4.1	757	8	US-60-685-372-466
12	156	4.1	888	6	US-10-450-763-44262
13	156	4.1	888	6	US-10-450-763-54269
14	150.5	4.0	1391	6	US-10-450-763-38863
15	149	3.9	1857	1	PCT-US03-10870-2173
16	149	3.9	1857	7	US-11-009-554-21
17	148.5	3.9	1192	7	US-11-097-143-5718
18	148	3.9	1208	6	US-10-450-763-54526
19	147.5	3.9	951	8	US-60-685-372-462
20	147	3.9	3394	1	PCT-US03-10870-529
21	145.5	3.8	899	1	PCT-US04-37982-1140
22	145	3.8	1606	1	PCT-US05-18850-981
23	144.5	3.8	1582	6	US-10-535-571-39
24	143.5	3.8	1413	7	US-11-097-143-12927
25	142.5	3.7	1572	6	US-10-990-328A-11742
26	142	3.7	141.5	27	US-10-330-773A-904
27	141.5	3.7	141.5	7	US-11-051-720-1563
28	141.5	3.7	1521	7	US-11-051-720-1564
29	141	3.7	1911	7	US-11-097-143-9906
30	140.5	3.7	1714	6	US-10-450-763-32497
31	139.5	3.7	635	6	US-10-450-763-52460
32	139	3.7	1346	7	US-11-060-005-2
33	138.5	3.6	1714	6	US-10-450-763-36776
34	138.5	3.6	1976	7	US-11-097-143-35049
35	138	3.6	759	7	US-11-097-143-18597
36	138	3.6	1142	7	US-11-044-051-73
37	138	3.6	1155	6	US-10-450-763-54525
38	137.5	3.6	825	8	US-60-710-726-1424
39	137.5	3.6	1416	7	US-11-097-143-23289
40	137.5	3.6	3067	7	US-11-097-143-34878
41	137	3.6	1203	1	PCT-US03-10870-2174
42	137	3.6	1715	7	US-11-097-143-1059
43	136.5	3.6	532	6	US-10-950-095-28
44	136.5	3.6	1543	7	US-11-097-143-34911
45	136	3.6	904	6	US-10-450-763-52462
46	136	3.6	1049	6	US-10-540-634-8
47	136	3.6	1235	6	US-10-540-634-6
48	135.5	3.6	1735	1	PCT-US05-17105-6441
49	135.5	3.6	1780	7	US-11-097-143-41241
50	134.5	3.5	314	6	US-10-914-020-4685
51	134.5	3.5	1253	7	US-11-185-924-18
52	134.5	3.5	1253	8	US-60-710-726-546
53	134.5	3.5	1259	6	US-10-940-774A-10366
54	134.5	3.5	2342	8	US-60-664-936-1036
55	134.5	3.5	2342	8	US-60-664-936-1038
56	134.5	3.5	2468	1	PCT-US05-10454-230
57	134.5	3.5	2468	8	US-60-664-936-1039
58	134.5	3.5	2468	8	US-60-664-936-1046
59	134.5	3.5	2468	8	US-60-664-936-1047
60	134.5	3.5	2519	6	US-10-450-763-46995
61	134.5	3.5	2522	6	US-10-940-774A-10237
62	134.5	3.5	2622	6	US-10-450-763-36777
63	134	3.5	1049	7	US-11-051-720-1565
64	134	3.5	1439	7	US-11-124-368A-291
65	134	3.5	1526	7	US-11-051-720-1566
66	134	3.5	3012	8	US-60-710-726-1384
67	134	3.5	5002	7	US-11-097-143-17961
68	133.5	3.5	748	6	US-10-330-773A-394
69	133.5	3.5	896	1	PCT-US04-17965-2159
70	133.5	3.5	896	1	PCT-US04-17965B-2159
71	133.5	3.5	896	1	PCT-US04-17965C-2159
72	133.5	3.5	896	6	US-10-863-905-2159
73	133.5	3.5	1086	8	US-60-643-717-3535
74	133.5	3.5	1192	1	PCT-US05-10211-3
75	133.5	3.5	1192	7	US-11-090-836-3
76	133.5	3.5	1192	7	US-11-090-846-3
77	133.5	3.5	1192	7	US-11-090-847-3
78	133.5	3.5	1669	6	US-10-330-773A-392
79	133.5	3.5	2244	1	PCT-US05-18850-765
80	133	3.5	513	6	US-10-940-774A-6452
81	133	3.5	513	7	US-11-185-924-16
82	133	3.5	534	6	US-10-940-774A-10364
83	133	3.5	534	6	US-10-489-448-1881
84	133	3.5	1439	6	US-10-450-763-40959
85	132.5	3.5	2805	8	US-60-664-936-1044
86	132.5	3.5	2817	1	PCT-US04-37982-299
87	132.5	3.5	2870	1	PCT-US04-37982-297
88	132.5	3.5	2871	1	PCT-US04-37982-293
89	132.5	3.5	2871	6	US-10-471-571A-2922
90	132	3.5	974	6	US-10-990-328A-12859
91	132	3.5	2753	6	US-10-940-774A-7659
92	132	3.5	2753	6	US-10-940-774A-7660
93	132	3.5	372	6	US-10-914-020-5746
94	131.5	3.5	509	7	US-11-097-143-33774
95	131.5	3.5	808	1	PCT-US03-10870-722
96	131.5	3.5	2858	7	US-11-097-143-40242
97	131.5	3.5	3060	7	US-11-097-143-984
98	131.5	3.5	3060	7	US-11-097-143-984

Sequence 904, App  
Sequence 1563, Ap  
Sequence 1564, Ap  
Sequence 9306, Ap  
Sequence 32497, A  
Sequence 52460, A  
Sequence 2, Appli  
Sequence 36776, A  
Sequence 35049, A  
Sequence 18597, A  
Sequence 73, Appl  
Sequence 54525, A  
Sequence 1424, Ap  
Sequence 23289, A  
Sequence 34878, A  
Sequence 2174, Ap  
Sequence 1059, Ap  
Sequence 28, Appl  
Sequence 34911, A  
Sequence 52462, A  
Sequence 8, Appli  
Sequence 6, Appli  
Sequence 6441, Ap  
Sequence 41241, A  
Sequence 4685, Ap  
Sequence 18, Appl  
Sequence 546, App  
Sequence 10366, A  
Sequence 1036, Ap  
Sequence 1038, Ap  
Sequence 230, App  
Sequence 1039, Ap  
Sequence 1046, Ap  
Sequence 1047, Ap  
Sequence 46995, A  
Sequence 10237, A  
Sequence 36777, A  
Sequence 1565, Ap  
Sequence 291, App  
Sequence 1566, Ap  
Sequence 1384, Ap  
Sequence 17961, A  
Sequence 394, App  
Sequence 2159, Ap  
Sequence 2159, Ap  
Sequence 2159, Ap  
Sequence 2159, Ap  
Sequence 3335, Ap  
Sequence 3, Appli  
Sequence 3, Appli  
Sequence 3, Appli  
Sequence 3, Appli  
Sequence 392, App  
Sequence 765, App  
Sequence 6492, Ap  
Sequence 16, Appl  
Sequence 548, App  
Sequence 10364, A  
Sequence 1081, Ap  
Sequence 40959, A  
Sequence 1044, Ap  
Sequence 299, App  
Sequence 237, App  
Sequence 293, App  
Sequence 232, Ap  
Sequence 12859, A  
Sequence 7659, Ap  
Sequence 7660, Ap  
Sequence 5746, Ap  
Sequence 33774, A  
Sequence 722, App  
Sequence 40242, A  
Sequence 984, App



```
Db 382 EDKEDDSEALGEAES-----DHPSPH-DQRAHFRGWGVRPGKETEESAEADWG 431
Qy 588 SKTPSESVIAISECHTLLSCVKQLGSG-----ESECPSQVORDVLSGGRHHTH----- 635
Db 432 EAEPCPRVAI-----YVPGKPPPPAPPRLRLRLKRLKRPETPHDDPPET 480
Qy 636 -VREKVTFLVEEVTYYIS-----GDREKGPWEEFARDGCRFOKRIQETEDAIGYCLTF 689
Db 481 PLKARKVRPSEKVTVHFLAVWAGPAQAARQGPWELARDRSFARRITQAEELSPCLTP 540
Qy 690 EHRERMPNRLOGTCFKGLNLVKQ 712
Db 541 AARARAWARLNPPLAPALPTQ 563

RESULT 3
US-11-060-005-4
; Sequence 4, Application US/11060005
; GENERAL INFORMATION:
; APPLICANT: Irwin H. Gelman
; TITLE OF INVENTION: TUMOR SUPPRESSOR GENE
; FILE REFERENCE: A30558-A-FWC-A- 070156 0597
; CURRENT APPLICATION NUMBER: US/11/060,005
; CURRENT FILING DATE: 2005-02-17
; PRIOR APPLICATION NUMBER: 09/902,432
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 08/978,277
; PRIOR FILING DATE: 1997-11-25
; PRIOR APPLICATION NUMBER: 08/665,401
; PRIOR FILING DATE: 1996-06-18
; PRIOR APPLICATION NUMBER: 08/635,121
; PRIOR FILING DATE: 1996-04-19
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 1596
; TYPE: PRT
; ORGANISM: Rattus norvegicus
US-11-060-005-4
```

```
Query Match 4.4%; Score 167; DB 7; Length 1596;
Best Local Similarity 19.8%; Pred. No. 0.58;
Matches 151; Conservative 112; Mismatches 286; Indels 214; Gaps 35;

Qy 6 GGSRKRL-----GPR-----AGRFWPPPPRRSQAGSSKPT 38
Db 673 GSSKKRARKASSDDEGPRTLGDSHRAEASKDXEAGTDVAPSTQEQDQAGSSPE 732
Qy 39 PLGPENSGNPTLLSSAQPETRVSYWTKLLSOLLAPLPCLLQV-----LIWSQLFGGMF 92
Db 733 PAGSPSEGE-----GVSTW-SFKELVTPRKKSKEBEKAEDSSVEQLSTEIE 780
Qy 93 PTR---WLDPAGVYALRALKGRBKPAAPTAQKSL-----SLQDSSDPSVTS--PIDW 142
Db 781 PSREESWV-----SIKKFIPGRKRGADGKEQATVEDSGPVEINEDDPNPAVPLS- 833
Qy 143 LBEIGHQYSPDPLKELKAGSALDPAQAFLLEQLQWGLVLLPSSLQSLYNSRELGS 202
Db 834 -----EYNABV-REKMEAGNTELP-----QLLGA-----VYVSEELSK 866
Qy 203 SPSPGLNIQRIDNFSVVSYL--LNPSYLDGF-----PRLEVSQYSDGNSE 246
Db 867 TLVHTVSVAVIDGTTRAVTSVEERSPSMISASVTPELHTAGEAMPPEVTEKDIIAET 926
Qy 247 VVGFTQITPSSCLREHCHPQ-PLSABLIPASWQGGCPPLSTEGELPEIHLHMRKRLFLQ 305
Db 927 PVLQTLL-PEGKADHDWVTSEVDFTSEAVTAT-ETSEALRTEEVTEASGAE-ETTDVVS 983
Qy 306 QANKQDLP-----TPDQ--DNGVHSLREE-----HSLLRMDPKHCRDN----PTQVPA 349
Db 984 AVSQTLDSPDPTTERTATPVQEVESGVLDTTEEBERTQAILQAVADKVKESQVPATQTVQR 1043
```

```
Qy 350 AGDIPGNTQESTEKIELLTTEVPLALAEESPSGCPSPSEIPMEKPGEGRISVVDYSYL 409
Db 1044 TGSKALEKEVEEVEDSEVLASEKEKDVMPKGPVQEAEGAHLAQSGSETQATPESLEVPEV 1103
Qy 410 EGDLPISARAPACS-----NKLDIYILGASSDLETSSDPEG-----EDWDEAEODGFD 458
Db 1104 TADVDHVA--TCOVIKLQQLMEQAVAPSESETLTDSETNGSTPLASDRTADGTQOQDETID 1161
Qy 459 SDSLSDSLQEQDPEGLHLWNFSVDPYNPQNFTATICTAARIVPPEPS-----D 509
Db 1162 SQDSKATAVRQS-----QVTEBEATAQKEEPTLNNVPAQE 1200
Qy 510 SEKDSLKGSLENSQ---SGSLPETPEHSSGED---DM---ESSADEABSLKLMNSFCN 561
Db 1201 EHGEPEGRDVLEPTQOQELTAAAVPVLAKEVVGQEGEVDMLDGKVKEEQE----- 1250
Qy 562 SDDPYNPLNFKAPQOTSGENKGCGRDKTPSEISVAISECHTLLSCVKQLLGSQSECPD 621
Db 1251 -----VFVHSGPNSQKAADVTDSE--VMGVAGCQEKESKEVQSLSLEEGEMET 1297
Qy 622 SVQRDVLSSGRHTHVKRKVKVTFLEEVTEYYISGDEDRKGPWEE 664
Db 1298 DVE-----KEKRETKEQVSE---EGEQETAAPHE 1325

RESULT 4
US-10-450-763-57852
; Sequence 57852, Application US/10450763
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 790CIP3/US
; CURRENT APPLICATION NUMBER: US/10/450,763
; CURRENT FILING DATE: 2003-06-11
; PRIOR APPLICATION NUMBER: PCT/US01/08631
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/649,167
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 60736
; SOFTWARE: Custom
; SEQ ID NO 57852
; LENGTH: 1562
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (1524)..(1556)
; OTHER INFORMATION: HECT-domain (ubiquitin-transferase) domain identified by
; OTHER INFORMATION: eMATRIX, accession number PF00632C, p-value=3.302e-23, raw score
; OTHER INFORMATION: 20.66
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (1257)..(1562)
; OTHER INFORMATION: HECT-domain (ubiquitin-transferase) domain identified by
; OTHER INFORMATION: Pfam, accession name HECT, E-value=4.3e-127, Pfam score of 435.6
US-10-450-763-57852

Query Match 4.2%; Score 160.5; DB 6; Length 1562;
Best Local Similarity 20.9%; Pred. No. 1.3;
Matches 161; Conservative 99; Mismatches 277; Indels 235; Gaps 40;

Qy 31 AGSSKEPTP-LGPENSGNPTLLSSAQPETRVSYWTKLLSOLLAPLPGL-LQKVLIWSQL 87
Db 123 SGALRATTSVTVKNNSAAPIFKSGIAGDETVOGGGSRRLISFSLSDFQAMGLKK----- 175
Qy 88 FGMGF--PTRWLDFA---GVYSALRAL--KGREKPAAPTAQKSLSLQDSSDPSVTSPL 140
Db 176 --GMFNPDPYLKISIQPGKHSIFPALPHGQER---RSKIIG-----NTWNPI 219
Qy 141 DWLEEGIHQYSPDP-LKLELKAKGSGALDPAQAQF-----LLEQQLMWGLVLLPSS 189
```

```
Db 220 -WQAEQSFVSLPTDVLIEVVDKFAKGRPIIKRFLGKLSMPVQRLRHAIGDRVVSYT 278
Qy 190 LOSRLYSNRELG-----SSPGPLNIQRIIDNFSVVSYLLNPSYLDLC 230
Db 279 LGRRLPTDHSVQLOFRFEITSSIHDPDEEISLSTEPESAQIQD--SPMNNLMESGSGE- 335
Qy 231 PPRLEVSQNSDGNSEVGVGFTLTPESSCLREDHCHPQPLSAELIPASWQGPCPLSTEGL 290
Db 336 -PRSEAPESSEWKPEQLGEGSVDPGPNQSIELSRPAEEAAVITEAGDQGMVSGPEGA 394
Qy 291 PIHHLRMKRLFLQOANKG-QDLPTPDQDNGYHSLREHSLLRMDPKHCRDNPQTQFVPA 349
Db 395 GEL-----LAQVKDIQAPASABELAQDLGEEASALLLED----- 431
Qy 350 AGDIPGNTQESTEEKIELLTTEVPLALEE---ESPSEGCPSSEIPMEKEPGEGRISVDY 406
Db 432 -GEAPASTKEEPL-EEATTQSRAGREBEKEQEEBEGDVST-----LEQEGEGLQ---- 479
Qy 407 SYLEGDLPIASRPACSNKL--IDYILGGASSDLET-----SSDPEGEDWDDEA 452
Db 480 --LRASVKRKSRP-CSLPVSELETVIASACGDPETPRTHYIRIHTLLHSMPSAQGSAE 536
Qy 453 EDGFDSDSSLSDSLQDPEGLHLWNSFCSDVPYNPNFTATTQTAARIIVPEEPSDEK 512
Db 537 EEDGAEESTLKDS---SEKQGL-----SEVDIVA---ADPSALEE 571
Qy 513 DLSGKSDLENSQSGLPET--PEHSSG-----EEDWESSADAESLKLWNSF 559
Db 572 D-----REEPEGATPGTAHPGSHGHPFPLSLANGAAQDGTHTPSTGESDS----- 616
Qy 560 CNSDDPNPLNFKAPQTSGENEKGRDSTKTPSEIVASISCHTLLSKVOLLGQSEK 619
Db 617 -----SPRQGDHSCGEC-DASCCSPSCYS-SSCYS-TSCYSSSCYSASCYS 660
Qy 620 PDSVORDVLSGGR---HTHVKKRKVTFLFVETVYISGDEDRK----- 659
Db 661 P-----SCYGNRPFASHTRFSSVDSAKISESTVFSSQDDEEENSAPESVPSMQSPELD 715
Qy 660 -----GPW-BEFARDGCRFKRIQETEDAIGYCLTFEHRMFMNRLQGTG 703
Db 716 PESTNGAGPWQDELAAPSGHVERSPGLESPVA-----GPSNRREGEC 758

RESULT 5
US-10-450-763-38864
; Sequence 38864, Application US/10450763
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 790CIP3/US
; CURRENT APPLICATION NUMBER: US/10/450,763
; CURRENT FILING DATE: 2003-06-11
; PRIOR APPLICATION NUMBER: PCT/US01/08631
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/649,167
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 60736
; SOFTWARE: Custom
; SEQ ID NO 38864
; LENGTH: 1702
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (1470)..(1498)
; OTHER INFORMATION: HECT-domain (ubiquitin-transferase) domain identified by
; OTHER INFORMATION: eMATRIX, accession number PF00632B, p-value=3.700e-21, raw score
; OTHER INFORMATION: 18.45
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (1257)..(1628)
```

```
; OTHER INFORMATION: HECT-domain (ubiquitin-transferase) domain identified by
; OTHER INFORMATION: Pfam, accession name HECT, E-values=5.5e-80, Pfam score of 279.1
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(1702)
; OTHER INFORMATION: Xaa = X or * as defined in Table 2
US-10-450-763-38864
```

Query Match 4.2%; Score 160.5; DB 6; Length 1702;  
Best Local Similarity 20.9%; Pred. No. 1.4;  
Matches 161; Conservative 99; Mismatches 277; Indels 235; Gaps 40;

```
Qy 31 AGSSKFFTP-LGPNSEGNPTLLSSAQPETRVSVYWT-KLLSQLLAPLPGL-LQKVLWSQL 87
Db 123 SGALRAITPSTVTKNSAAPFKSIGADETVQGGSRRLISFSLDSDFQAMGLK----- 175
Qy 88 FGMFNP--PTRWLDP-----GVYSALRAL--KGRKPAAPATAKSLSSLOLSDSDSVTSPL 140
Db 176 --GMFFNPDPYLKISIQGKHSIFPALPHHQBGR-----RSKIIG-----NTVNP 219
Qy 141 DWLEEGIHMQYSPPD-LKLLEKAKGSALDPAQAQF-----LLEQQLWGVELLPS 189
Db 220 -WQAEQSFVSLPTDVLIEVVDKFAKGRPIIKRFLGKLSMPVQRLRHAIGDRVVSYT 278
Qy 190 LOSRLYSNRELG-----SSPGPLNIQRIIDNFSVVSYLLNPSYLDLC 230
Db 279 LGRRLPTDHSVQLOFRFEITSSIHDPDEEISLSTEPESAQIQD--SPMNNLMESGSGE- 335
Qy 231 PPRLEVSQNSDGNSEVGVGFTLTPESSCLREDHCHPQPLSAELIPASWQGPCPLSTEGL 290
Db 336 -PRSEAPESSEWKPEQLGEGSVDPGPNQSIELSRPAEEAAVITEAGDQGMVSGPEGA 394
Qy 291 PIHHLRMKRLFLQOANKG-QDLPTPDQDNGYHSLREHSLLRMDPKHCRDNPQTQFVPA 349
Db 395 GEL-----LAQVKDIQAPASABELAQDLGEEASALLLED----- 431
Qy 350 AGDIPGNTQESTEEKIELLTTEVPLALEE---ESPSEGCPSSEIPMEKEPGEGRISVDY 406
Db 432 -GEAPASTKEEPL-EEATTQSRAGREBEKEQEEBEGDVST-----LEQEGEGLQ---- 479
Qy 407 SYLEGDLPIASRPACSNKL--IDYILGGASSDLET-----SSDPEGEDWDDEA 452
Db 480 --LRASVKRKSRP-CSLPVSELETVIASACGDPETPRTHYIRIHTLLHSMPSAQGSAE 536
Qy 453 EDGFDSDSSLSDSLQDPEGLHLWNSFCSDVPYNPNFTATTQTAARIIVPEEPSDEK 512
Db 537 EEDGAEESTLKDS---SEKQGL-----SEVDIVA---ADPSALEE 571
Qy 513 DLSGKSDLENSQSGLPET--PEHSSG-----EEDWESSADAESLKLWNSF 559
Db 572 D-----REEPEGATPGTAHPGSHGHPFPLSLANGAAQDGTHTPSTGESDS----- 616
Qy 560 CNSDDPNPLNFKAPQTSGENEKGRDSTKTPSEIVASISCHTLLSKVOLLGQSEK 619
Db 617 -----SPRQGDHSCGEC-DASCCSPSCYS-SSCYS-TSCYSSSCYSASCYS 660
Qy 620 PDSVORDVLSGGR---HTHVKKRKVTFLFVETVYISGDEDRK----- 659
Db 661 P-----SCYGNRPFASHTRFSSVDSAKISESTVFSSQDDEEENSAPESVPSMQSPELD 715
Qy 660 -----GPW-BEFARDGCRFKRIQETEDAIGYCLTFEHRMFMNRLQGTG 703
Db 716 PESTNGAGPWQDELAAPSGHVERSPGLESPVA-----GPSNRREGEC 758
```

## RESULT 6

```
US-10-990-328A-9765
; Sequence 9765, Application US/10990328A
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; TITLE OF INVENTION: POLYMORPHISMS IN NUCLEIC ACID MOLECULES
; TITLE OF INVENTION: ENCODING HUMAN ENZYME PROTEINS, METHODS OF DETECTION AND
; TITLE OF INVENTION: USES THEREOF
```

```
; FILE REFERENCE: CL001495
; CURRENT APPLICATION NUMBER: US/10/390,328A
; CURRENT FILING DATE: 2004-11-17
; NUMBER OF SEQ ID NOS: 558824
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9765
; LENGTH: 1585
; TYPE: PRN
; ORGANISM: Homo sapiens
US-10-990-328A-9765

Query Match          4.2%; Score 158.5; DB 6; Length 1585;
Best Local Similarity 20.9%; Pred. No. 1.7;
Matches 161; Conservative 99; Mismatches 277; Indels 235; Gaps 40;

Qy 31 AGSKFTPT-LGPNNGNPTLLSSAQPTRVSYWT-KLLSQLLAPLQL-LQVLINSQL 87
Db 146 SGALRATPTSVTVKNSAAPFKSGADETVCGQSRRLISFSLDFQAMGLK-----198
Qy 88 FGMF--PTRWLDPF---GVYSALRAL--KGREKPAAPTAQKSLSLQLDSSDPSTVPL 140
Db 199 --GMFFNPDPYLKISIQPKGHSIFPALPHHGOER-----RSKIG-----NTVNPI 242
Qy 141 DWLEEGHWHQVSPD-LKLEKAGSALDPAQAF-----LLEQQLWGLVLLPSS 189
Db 243 -WQAEQSFVSLPTDVLEIEVKDFPAKSRPIKRFGLKSLMPVQRLERHAIGDRVVSYT 301
Qy 190 LQSLRYNRELG-----SSPSGPNLIQRIQIDNFSVSYLLNPSYLD 230
Db 302 LGRLPDTHGGVQLQFPEITSSHPDDEISLSTEPESAQIQD--SPMNLMSGSGE- 358
Qy 231 PPRLEVSQNGSDGNSEVVGFOTLPSSCLREDHCPQLSAELIPASWQCCPLSTEGL 290
Db 359 -PRSEAPSESSWKPEQLGEGSVDPGQNSIELSRPAEERAAVITEAGDQGMVSGPEGA 417
Qy 291 PEIHLRMKLEFLQANKG-QDLPDQDNGYHSLLEHSLLRMDPKHCRNDNPTQVPA 349
Db 418 GEL-----LAQVQKIQAPASAEALAEQLDLGEASALLLED-----454
Qy 350 AGDIPGNTOSTEKIELLTTEVPLALEE---ESPSEGCPSSEIPMEKEPGEGRISVVDY 406
Db 455 -GEAPATKBEPLR--EATQTSRAGREEBEKEQEEEDVST-----LEQEGRLQ----502
Qy 407 SYLEGDLPISARPACSNKL--IDYILGASSDLET-----SSDPEGDWDEEA 452
Db 503 --LRASVKRKSRL--CSLPVSELETVIASACGDPETPRTHYIRIHTLLHSPSAQGGSAE 559
Qy 453 EDDGFDSSLSLSDQLDQDPGLHWNFSFVDPYFNQNFATQTAARIVPEPSPSEK 512
Db 560 EEDGAEESTLKDS---SEKQGL-----SEYDTVA---ADPSALEE 594
Qy 513 DLGSKSLDSSQSGSLPET--PEHSSG-----EEDDWESSADAESLKNSP 559
Db 595 D-----REEPGATPGTAHGHSGHPFSLANGAAQDGTPTSGSEDS-----639
Qy 560 CNSDDPYNPLNFKAPFQTSGENEGCRDSTKTPSESIIVASISCHTLLSKVQLLGSQSE 619
Db 640 -----SPRQGDHSCGCG-DASCCSPCY-S-CY-S-TCYSSSCYASCY 683
Qy 620 PDSVORDVLSGR---HTHVKKVTFLEEVYIISGDEDRK-----659
Db 684 P-----SCYNGNRFASTRFSSVDSAKISESTVFSSODDEEENSAFESVPDSMQSPELD 738
Qy 660 -----GPW--EFPARDCCRQKRIQETEDAIQYCLTFEHRERMFNRLLQGC 703
Db 739 PESTNGAGPQWDELAAAPSGHVSPGLESPVA-----GPSNRREGEC 781

Query Match          4.2%; Score 158.5; DB 7; Length 5303;
Best Local Similarity 20.4%; Pred. No. 7.5;
Matches 93; Conservative 56; Mismatches 159; Indels 147; Gaps 16;

Qy 309 KGQDLPTDQDNGYHSLLEHSLLRMDPKHCRNDNPTQVPAAGDIPGNTOSTEKIELL 368
Db 4447 KMQDVEKPEDDSGSEEEEDLNKMG-----ETEGAEKL 4482
Qy 369 TTEVPLALREESPSGCPSEIPM-EKEPGEGRISVVVSYLSEGLDLPISARPACSNKLID 427
Db 4483 DDQIWDGDEEKPEE---EQEPDMEEDQKGSKDEK-----4517
Qy 428 YILGASSDLETSSDPEGDWDEEAEDGFDPSLSLSDS-----467
Db 4518 -----AHNDLDTKNDAKEDGKDEHEKQGLDATNEPSGEDKKEQAKIDDMKDPMDDEE 4572
Qy 468 -----LEQDPEGLHL-WNSFCSDV-----PYNPQNFTATIQTAARIVP 504
Db 4573 QTNAMNELEPEPEPEEMDLGMNVNDEGHDDQDQDPTDENPFDIDAMKENMQPAEB--P 4630
Qy 505 EEPDSGSEKOLS-----GKSDLENSQSGSLPETPEHSSGDEDDWESSADAESLKLW 556
Db 4631 EADGDDHDANESGDPQSDGSDSEDEEAGTEAKPAEDHGEGEATPEDEKDEATQK--4688
Qy 557 NSFCNSDDPYNPLNFKAPFQTSGENEK-----GCRDSTKTPSESIIVASISCHTLLS 606
Db 4689 RGELEDEDDSKPED--SPEDSKEEKEKEKEPEEHSQSKDKASKENQVSMPE-----4740
Qy 607 CKVQLLGSQSECPDSVQRDVLSGGRHTHVKKVTFLEEVYIISGDEDRKGPWEPA 666
Db 4741 -----TDQSSADQVQBP-----QDPDIKQD-----QKLDQETGEKGGVQAENDA 4783
Qy 667 RDGCRFQKRIQETEDAIQYCLTFEHRERMFNRLLQGC 701
Db 4784 DDG--GHQVAFQETVVSQEDRKNERQTQEKRKQG 4816

RESULT 8
US-11-097-143-31983
; Sequence 31983, Application US/11097143
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
```

```

Db      1906 APVDSDEDTAPSDEKIPSVSGEVPETVTITASPQAAREDELKTAPSESPS----- 195
Qy      561 NSDDPYNPLNFKAFFQTSGENEGCRDSTKPSSIVAISECHTLGCKY 609
Db      1958 -STDKYPETEYQXP-----EDTKRADETPT-ESVTQVSVDATSTSAPV 1997


RESULT 9
US-10-877-346-62
; Sequence 62, Application US/10877346
; GENERAL INFORMATION:
; APPLICANT: Gerlach, Valerie L
; APPLICANT: MacDougall, John R
; APPLICANT: Smithson, Glennda
; APPLICANT: Millet, Isabelle
; APPLICANT: Stone, David
; APPLICANT: Gunther, Erik
; APPLICANT: Ellerman, Karen
; APPLICANT: Grosse, William M
; APPLICANT: Alsobrook II, John P
; APPLICANT: Lepley, Denise M
; APPLICANT: Burgess, Catherine E
; APPLICANT: Padigar, Muralidhara
; APPLICANT: Kekuda, Rameshly A
; APPLICANT: Spytek, Kimberly A
; APPLICANT: Leach, Martin D
; APPLICANT: Shimkets, Richard A
; TITLE OF INVENTION: Novel Proteins and Nucleic Acids Encoding Same
; FILE OF INVENTION: 21402-124
; CURRENT APPLICATION NUMBER: US/10/877,346
; CURRENT FILING DATE: 2004-06-25
; PRIOR APPLICATION NUMBER: US/09/964,956
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: 60/235,631
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: 60/235,633
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: 60/235,808
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: 60/236,064
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: 60/236,065
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: 60/236,066
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: 60/236,135
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: 60/237,434
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/238,321
; PRIOR FILING DATE: 2000-10-05
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 127
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 62
; LENGTH: 2766
; TYPE: PRF
; ORGANISM: Rattus norvegicus
US-10-877-346-62


Query Match          4.1%; Score 157; DB 6; Length 2766;
Best Local Similarity 19.1%; Pred.No. 4.1;
Matches 171; Conservative 105; Mismatches 261; Indels 360; Gaps 47;

Qy      3   PGTG--GSRKGLGPAGFRFWPPFPFRSQAGSKKEFTPLGPSNGNPILLSSA-QPETR 59
        |||||    | : | : | : | : | : | : | : | : | : | : | : | : | : 
Db      1164 PGTGWGDGSEHL-----CSPGKSr---EVHPDSSSTPTVAEQVHOPES- 1203
                : | : | : | : | : | : | : | : | : | : | : | : | : 

Qy      60 VSYWTKLLQLLAPLGCLLKVLWSQLFGGMPTTRWLDFAGVYSALRALKGREKPAAPT 119
        |||||    | : | : | : | : | : | : | : | : | : | : | : | : | : 
Db      1204 -----LSQPVSPTSFE-----SQGISKKMP-----PSQRCVSPREKASTTP 1241
                : | : | : | : | : | : | : | : | : | : | : | : | : 

Qy      120 AQSLSSLQDSDDPSTGPLDWLEGIHQVSPDDLKELKAKGSALDPAA--QAFLL 177
```

```
Db 1242 -----DSSRA-----WAAPG-----DSSPTRRIAVPMGTGAAPATAIPQASLVS 1281
Qy 178 QQ-----LWVGL-LPSSIQ-----SRLYSNRELSSPSGGLNIQR----- 212
Db 1282 QERSGLSGPSKGLGTKELCIPKSLKOGALLEDAPASGKMSHASSPSGVPATERTLSGS 1341
Qy 213 -----IDNFSVSYLLNPYSYLCDFRLEVSQNSDGNSEVGFQTLTPSSCLREDHC 265
Db 1342 PENPVTIDNFEIA-----SEARLSQPKADCRAGHDTFESQPGAGAGSSSH- 1391
Qy 266 HPOPLSAELIPA-----SNQGCCP-----LSTEGLEPIH-----HIRMKLEFLOQANK----- 309
Db 1392 HAQWRSQTSRPTGTGTGTPPPQWALQPSVLDSHPDKHLAVNK-TFLNYSRNFSN 1450
Qy 310 -----GQDLPTDQNGVHSLSEEGHSLLRMDPKCRDNP-----TQFVP 348
Db 1451 FHEDSISLSPGGSSSESP-SSMYGNVEDSSS-----DPESLAEDPFGAARNWSPPLSP 1504
Qy 349 AAGDIPGNTQSTEEKIELLTTEVPLALAEBSSEGCPSSEI-----PMEX 394
Db 1505 ESSPKEGSS-ESDERIEICST-----DGCFTPTVTPPTQVALCPVLPVQ 1551
Qy 395 E-----PGEGRISVDYSYLEGLDLPISAR-----PACSN 423
Db 1552 RAVCKPVDICERACFVPGASRTSIPDSSQPSFLDVSSEBPETWASINASQNHMPVCTE 1611
Qy 424 KLIDVI-----LGAGSSDLETS-----SDPE-----CED-WDEEAEDDGF----- 457
Db 1612 GIMDVTSSNMGDSQSSQWTRHCRNAPFVLGNFDMVNDLGRDLLDEGAPKEGAAAASVM 1671
Qy 458 -----DSDSLSLD-----SDLEQDPEGLHL 477
Db 1672 RSVFALGAEKPKGEAVLADLHIAERGNLELLQPKTISRPPILTWFKEINKDSQSHL 1731
Qy 478 -----WNSFCSVDPNPQNFATITQTAARIVPEEP-----SDSEKDLCKSLDN----- 522
Db 1732 RSTSEKQSMALALGPGSKAMVNTGHRKGVTVPKSPSRQSKQENKDLPPKSPVETLGN 1791
Qy 523 -----SSQSGSLPETP-----HSSGEEDDWESSADEAE-SLKLNWS-----F 559
Db 1792 CQKPKCSPKLKRLNSKGAKEVPAVSTKSRNDRHRTLPSPQASHKMFKAVSHRLHI 1851
Qy 560 CNSDDPNPLNFKAPFOTSGENEGCR-----DSKTPSEISVAISECHTLCKVQLLSQOE 616
Db 1852 ADQEEPKN-----TAGDTPKPCQVPSKPPQAA-----LGLSLR 1885
Qy 617 SECPDSVQDVLSS-----GGRHTVHKRKVTFLEEVTEYVISGDEDRKGP 661
Db 1886 TSASDTSIRTPTSLTSPKLLPEQANSRFFM-----AVLESdTSCPTTSRPSRGP 1938
```

RESULT 10  
PCT-US05-18850-797

```
; Sequence 797, Application PC/TUS0518850
; GENERAL INFORMATION:
; APPLICANT: CHIN, Lynda
; TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF
; TITLE OF INVENTION: CANCER
; FILE REFERENCE: DFN-059-2
; CURRENT APPLICATION NUMBER: PCT/US05/18850
; CURRENT FILING DATE: 2005-05-27
; NUMBER OF SEQ ID NOS: 1528
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 797
; LENGTH: 1503
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US05-18850-797
```

Query Match 4.1%; Score 156.5; DB 1; Length 1503;  
Best Local Similarity 20.4%; Pred. NO. 2.1;

```
Matches 147; Conservative 82; Mismatches 249; Indels 243; Gaps 33;
Qy 26 PRSQAGSKFPTLPGPENGNTLSSAQPTFRVSYWTKLQLQLLAPLGLLQKVLWS 85
Db 826 PRRVP-----PDSLPTQGETQFTCLDIVPE-----DCLHQDISP-----DAVTPV 867
Qy 86 QLFQGMPTRWLPAGVYSALRAKLGREKPAATAQKSLSSQLDSSDPVTSPLDWLEE 145
Db 868 EILSTDARTHSLD-----NRQDSFG-----ESEETLRLTESDSVLAD--DILAS 910
Qy 146 GIHWQSPPLKLELAKGKAGSALDPAQAFLLEQOLGVELLPSSLSQSRYSNRELSSPS 205
Db 911 RVSUGSLPLGQELHKNKPFSEHHSHR-RLEKNLEAVETLQ-----LNSKDAKEA 962
Qy 206 GPLNIQRIDNFSVSYLLNPYSYLCDFRLEVSQNSDGNSEV----- 247
Db 963 GLVSALSSDSTQS-QS-LLEDLSAPFPASEPSLETDPDSLESVDVHEALLDSLSHTPKL 1021
Qy 248 -----VGFQTLTPSSCLREDHCHPQL-SABLIPAS-----WQGCCP-----LSTE 288
Db 1022 VPPDKPADSGYETENLESP-----EWTLLHAPAGTADSEPATTDGCGHSGLPNPVIVISDA 1078
Qy 289 GLPEIHHLRMKRLEFLQANKGODLPTD-----ODNGVHSLSEEHSLLRMDPKH 338
Db 1079 G-----DGRGTEV-TPETFTAGSQSYRDSAYFS----- 1107
Qy 339 CRDNPTQFVPAAGDIPGNTQSTEEKIELLTTEVPL-----ALAEBSPS--EGCPSSEIPME 393
Db 1108 --DNDSEPKRSEVPG-----TSPSALVLVQQLPEPVLPEQSPAAQDSCLEAR--K 1157
Qy 394 KBQEGRISVDYSYLEGLDLPISARPACSNKLDIYLLGGASSDLETSSDPEGEDW----- 448
Db 1158 SQPDESCLSLAHNS--SDLELATPEPAQT-----GVPPQVHPTEDEASSPWSVLNA 1207
Qy 449 -----DEEAEDD-----GFDSDSLSS-----DSLEQDPEGLHLNWSFCSDVPYN 488
Db 1208 ELSGDDFTQDRPCTLASTGNTNELLAYTNALSLSHSEG----- 1253
Qy 489 PONTATITQTAARIVPEEPSDESKDLSGKSDLNSQSGLSPETPEHSSGEEDDWESSAD 548
Db 1254 -----PKLEPDIEGKY-LGKLGVSGLMDLSDGMDADEED-ENSDD 1293
Qy 549 EASLKLWNSFCNSDDPNPLNFKAPFOTSGENEGKCRDSTPSESIVALSECHTLLSCK 608
Db 1294 SDEDLAFNLHLSSESEDETEHPVPIILSNEDGRHLR----- 1331
Qy 609 VQLLSQSESCPSVQDVLSSGGRHTVHKRKVTFLEEVTEYVISGDEDRKGPWEPEARD 668
Db 1332 -SLLKPTAANAPDLPED-----WKEKKAATFFDDVTYYLF-----DOETPTKELGPC 1379
Qy 669 G 669
Db 1380 G 1380
```

RESULT 11

```
US-60-685-372-466
; Sequence 466, Application US/60685372
; GENERAL INFORMATION:
; APPLICANT: Belouchi, Abdelmajid
; APPLICANT: Raelson, John V
; APPLICANT: Bradley, Walter E
; APPLICANT: Paquin, Bruno
; APPLICANT: Nguyen-Huu, Quynh
; APPLICANT: Croteau, Pascal
; APPLICANT: Allard, Rene
; APPLICANT: Little, Randall D
; APPLICANT: Cousineau, Johanne
; APPLICANT: Erdewegh, Paul V
; APPLICANT: Segal, Jonathan
; TITLE OF INVENTION: Genemap of the Human Genes Associated with Psoriasis
; FILE REFERENCE: US9908-5005-PR
; CURRENT APPLICATION NUMBER: US/60/685,372
```



```
; CURRENT FILING DATE: 2005-05-31
; NUMBER OF SEQ ID NOS: 2738
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 466
; LENGTH: 757
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-685-372-466

Query Match
Best Local Similarity 4.1%; Score 156; DB 8; Length 757;
Matches 118; Conservative 90; Mismatches 218; Indels 170; Gaps 30;

QY 3 PGTGSRKRLGRAGRFWRPFRRSQAGSKFTPLGPENSGMPTLLSSAQBPETRVSY 62
DB 77 PGTLPST---PVTSF---PGIPDTLPFGSAPLEAPMTPTVDDSPQ-----115
QY 63 WTKLLSQLLAPLGLLQKVLINSLQFLGGMFFTR-----WLDPAFAG 101
DB 116 -KKMLGQKATPPPPPLLSGLKK--GSLLPSPRLVNESEMAVASGHLNSTGVLLVGG 171
QY 102 VYSALRALGKREK---AAPTAKSL-----SIQLDSSDPSTVSPDLWLEEGIHW 149
DB 172 VLPMHGGIEIQTPNTVAASPAESVSQATIVMMPALPAPSAAPAVST-----TESVA- 225
QY 150 QYSPDPLKLELKAKGSALDPAQAFLLEQOLMGVELLPSSLQSRLYSNRELGSPPGSLN 209
DB 226 PVSQPDNCVPMBAVG---DPHTVTVSMDSS--EISMIINSIKKECFRS-GVAEAPVG---276
QY 210 IQRIDNFSVSYLLNPSVLDPCPRLE--VSYQNSDGNSEVVGFTLTPESSCLREDHC--265
DB 277 -----SKAPSIDGKELDLAEKMDIAVSYTGEBLDPETVG-----DIITAIIDKVDD 323
QY 266 HPQPLSAELIPASWQGC-----PPLSTEGLPETHHLRMKRLFLQOANKQDLPDP-----317
DB 324 HPEVLDAVAEAAALSCBENDDPQLPG-PWEHPIQ-----QERDPVPLPAPMTVK 375
QY 318 -----QNGYHSL---EEHSLRMDPKHCRDNPTQFVPAAGDIPGNTQSTE---362
DB 376 QERLDFEETENKGIHELVDIREPSAEIKVEPAEPVLSGAEIVAGVVPATSMPEPLRS 435
QY 363 -----EKI-----ELLTTEVPALAEESP-----SEGCPSSEIPEMEKEPGEGRISVVDYS 407
DB 436 QDLDEELGSTAAGRIEADVAIGKDETPLTNVKTASPSML-----SPSHGSPNPIED--489
QY 408 YLEGDLPI SARACSNKLIDYILGASSDLETS-SDPEGDWDEAEADDGDSDSLSD-465
DB 490 -----PLEAETQHKFMSDSLSKEESGTIFGQIKDAFGED-----EEDGVSEAAASLEEP 539
QY 466 -----SDLEQDPBGLHLWNSFCSDVPYNPQNTATTIQTAAARIVPEEPSPDSE 511
DB 540 KEEDQGEGLSEMDENPVSE-----SDDGFSIHNTLQSHTLADSLPSSPASSQ 589

RESULT 12
US-10-450-763-44262
; Sequence 44262, Application US/10450763
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 790CIP3/US
; CURRENT APPLICATION NUMBER: US/10/450,763
; CURRENT FILING DATE: 2003-06-11
; PRIOR APPLICATION NUMBER: PCT/US01/08631
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/649,167
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 60736
; SOFTWARE: Custom
; SEQ ID NO 44262
; LENGTH: 888
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-450-763-44269

Query Match
Best Local Similarity 4.1%; Score 156; DB 6; Length 888;
Matches 134; Conservative 64; Mismatches 209; Indels 176; Gaps 31;

QY 63 WTKLLSQLL-----APLPCLLQKVLINSLQFLGGMFFTRWLDPAFVYSALRALGKREK 115
DB 40 WPLLMMQLLMLLVKBAQFLWVKDPLQLTNPLGP--PEPWSSSHSHP-----RESP 90
QY 116 AAPTAQKSLSSQLDSSDPSTVSPDLWLEEGIHWQYSPDLKLE-----LKAAGSA-----166
DB 91 HAPTL--PADPWFDPHGLGPSASSEMP-----APQESTENLVFLDWDWDSAGELP 138
QY 167 LDPAQAFLLEQQLMGWVELLPSSLQSRLE--YSNRELGSPPGSLNLIQIRDNFSVSYLLNP 225
DB 139 LEP--EQFLASQQLDKDLSP--QERLPVSPKLLKDPARQWSLAEI--IGIIRQLSTP 191
QY 226 -----SYLD-----CPRLEVSQNSDGNSEVVG-----FOTLTPE-----256
DB 192 QSQKQTTQNEYSSTDTYPGSLPELKVSDPEPPGPPSEQVGPQFHLPEPONTETLEDI 251
QY 257 -SSCLREDHCHPQPLSAELIPASWQ----GCPLSTEGLPETHHLRMKRLFLQOANKQ 311
DB 252 QSSSLQQAEPALQPLLEEPSSMQEAPALPPES-----MESLTLPNHEV 298
QY 312 DLPTPDQONGYHSL-----EEHSLRMDPKHCRDNPTQFVPAAG 351
DB 299 SVQPPGEDQAYYHLPNITVKPADVEVTTSEATNETESSQAQ-----QETPIQF-PEEV 351
QY 352 DIPGNTQESTEEKIELLTTVPALAE-----SPSEGCPSSEI---PMEKE-----395
DB 352 E-PSATQQAEPAPPEPPMEHELISISEQQQPVQPS--TSREVSSTQETPQCPPEHH 408
QY 396 -----PGEGRISVVDYSYL-----EGDLPISARPAC--SNKLIDY-----ILGASSDLE 438
DB 409 EVTVSPPGHQTTHLDSFVSVKPPDVQLTIAAPSAEAVGTSLVQQAETTLRSGSGNDIE 468
QY 439 TSSDPEGDWDEAEADDGDSDSLSDLEQDPBGLHLWNSFCSDVPYNPQNTATTIQT 498
DB 469 PPAIQHG-----GPPLLSSESEAGPLAVQOETSFQSPSEFINNENPSPTQOE 515
QY 499 AARIVPEEPSPDKSLSGKSDLENSSQSGSLP-----ETPEHS 536
DB 516 AA---AEHPQTABEGESSLTHQEAQAPQTPFPNVAQPPSHS 555

RESULT 13
US-10-450-763-54269
; Sequence 54269, Application US/10450763
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 790CIP3/US
; CURRENT APPLICATION NUMBER: US/10/450,763
; CURRENT FILING DATE: 2003-06-11
; PRIOR APPLICATION NUMBER: PCT/US01/08631
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/649,167
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 60736
; SOFTWARE: Custom
; SEQ ID NO 54269
; LENGTH: 888
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-450-763-54269

Query Match
Best Local Similarity 4.1%; Score 156; DB 6; Length 888;
```

Best Local Similarity 23.0%; Pred. NO. 1.2;  
Matches 134; Conservative 64; Mismatches 209; Indels 176; Gaps 31;

Qy 63 WTKLLSQL-----APLEGLLQKVLVWSQLFGGMFTRWLDFAGVYSALRALKGREKP 115  
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |  
Db 40 WPLLMQALLWLVKEAQLPWLKPNLP--PEPMSSSHSHP-----RESP 90  
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |  
Qy 116 AAPTAKSLSSQLDSDSDSVTSPLDWLBEGHMQYSPDLKLE-----LKAKGSA---- 166  
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |  
Db 91 HAPTL--PADPMDFDHLGPSASSEMP-----APPQESTENLVFLDTWDSAGELP 138  
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |  
Qy 167 LDPAQAFLLEQLMGVELLPSSLQSRL--YSNRELSSPSGPLNQIRIDNSVVSYLLNP 225  
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |  
Db 139 LEP--SQFLASQDLKDKLSP---QERLPVSKKLKKPKPAQRWSLAEL--IGIIRQLSTP 191  
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |  
Qy 226 -----SYLD-----CFPRLEVSYQNSDGNSEVVG-----FQTITPE----- 256  
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |  
Db 192 QSQKOTLQNEYSSTDTPPGSLPPELRVKSDEFPGPSEQVPSQPHLPETONPETLEDI 251  
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |  
Qy 257 -SCLEDHCHPOPLSAELIPASWQ---GCPLSTEGLPEIHILRMKRLBFLQOANKQG 311  
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |  
Db 252 QSLSLQQEAQAPQLLEBEPSMQEADALPRESS-----MESLTLPNHEV 298  
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |  
Qy 312 DLPTPDODNGYHSL-----EEHSLLRMDPKHCEDNPTQFVPAAQ 351  
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |  
Db 299 SVQPGEQDAYHLPNITYKPADVEVITSEATNETESSAQ-----QETPIQP-PEEV 351  
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |  
Qy 352 DIPGNTQESTEEKIELTTEVPLALBEE---SPSEGCPSSBI---PMKE----- 395  
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |  
Db 352 E-PSATQQCAPIEPVPPMEHELSTEQOQVPQSE--TSREVESSTPQOETPGQPPHH 408  
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |  
Qy 396 -----PGEGRISVDVSYL-----EGDLPI SAR PAC--SNKLIDY-----ILGGASSDLE 438  
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |  
Db 409 EVTVSPRGHHTHLDSPSVVKPPDVQLTIAAEPSAEVGTSLVQOEATRILSGSGNDIE 468  
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |  
Qy 439 TSSDPGEDWDREAEDDFDSDSLSDLEQDDPEGLHLWNFSVCVDPYNQNFTATTQT 498  
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |  
Db 469 PPAIOHG-----GPPLLSESSEEAGPLAVQOETSFOSEPINNENPSPTQOE 515  
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |  
Qy 499 AARIVPEEPSDSKDI SGKSDLNSSQSGSLP-----ETPEHS 536  
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |  
Db 516 AA---AEHPOTABEGSSUTHQEARPAQTPEFFNVVVAQPPHS 555  
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |

RESULT 14  
US-10-450-763-38863  
Sequence 38863, Application US/10450763  
GENERAL INFORMATION:  
APPLICANT: Hyseq, Inc  
TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES  
FILE REFERENCE: 790CIP3/US  
CURRENT APPLICATION NUMBER: US/10/450,763  
CURRENT FILING DATE: 2003-06-11  
PRIOR APPLICATION NUMBER: PCT/US01/08631  
PRIOR FILING DATE: 2001-03-30  
PRIOR APPLICATION NUMBER: 09/540,217  
PRIOR FILING DATE: 2000-03-31  
PRIOR APPLICATION NUMBER: 09/649,167  
PRIOR FILING DATE: 2000-08-23  
NUMBER OF SEQ ID NOS: 60736  
SOFTWARE: Custom  
SEQ ID NO 38863  
LENGTH: 1391  
TYPE: PRP  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: DOMAIN  
LOCATION: (720)..(735)  
OTHER INFORMATION: WW/rsp5/WWP domain proteins domain identified by eMATRIX,  
accession number BL01159, p-value=8.568e-11, raw score of 13.85  
FEATURE:  
NAME/KEY: DOMAIN  
LOCATION: (705)..(989)

```
; ; OTHER INFORMATION: WW domain identified by Pfam, accession name WW, E-value=7.6e-10  
; ; OTHER INFORMATION: -26, Pfam score of 99.3  
US-10-450-763-38863
```

```
Query Match      4.0%; Score 150.5; DB 6; Length 1391;  
Best Local Similarity 19.8%; Pred. No. 4.1;  
Matches 137; Conservative 87; Mismatches 213; Indels 255; Gaps 36;
```

```
Qy   153 PPD-LKLELKAKGALDPAQAQ-----LLEOOLMGVELLPSSLOQRLLYSNRELG 201  
    |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::||  
Db   8 PTDLVEIVKDKFAKSPFIKRFLGKLMPVQLRHLERHAIGDRVSVYTLGRRLPTDH--- 64  
  
Qy   202 SSPSGPLNIQRIDNFVSUYNLLNSYLDCPFRLEVSYQNSDGNSEWVG--         ---QT 252  
    |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::||  
Db   65 --VSGQQF--FEITS-SIHFGYPN-----TTMLSSGRNDIFFFLPADDEISLS 110  
  
Qy   253 LTPESSCLRE-----DHCHPOPLSAELIP-----                276  
    |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::||  
Db   111 TEPESAIOQS PMNNLMESGBRPSEAPESSWKPEQLCEGSVPDPGNGQSIELSRA 170  
  
Qy   277 -----ASWGQCPISTGLPIHILMKRLFLOANKG-QDLPTFPDOONGVHSLEE 327  
    |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::||  
Db   171 EEAAVITEAGDGVMVSGPEGAGEL-----LAQVKDIQPAPSAEELAEQLDLGE 220  
  
Qy   328 EHSLLRMDPKHRDNPTQFVAPAGDIGNTQESTEEKIELLTTEVFPLEEE---ESPSEG 384  
    |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::||  
Db   221 EASALLED-----GEAFSTKEEFLE-EAATTSORAGREEEKQELEG 264  
  
Qy   385 CPSEIEMKEPGRGRI SVVDYSYLEGDLPISARPACSNKL--DYII LGASSLDLET--- 439  
    |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::||  
Db   265 DVST----LEQGEGRIQ-----LRASVKRSRP-CSLPVSELETVTIASCGDETPTRT 312  
  
Qy   440 -----SSDPREGEDWEAEADDFDSDSLSDSLEODPEGLHLWNFSFCSDPYNPQ 490  
    |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::||  
Db   313 HYIRHTLLSHMPSAQOGSAAEEDGAEEESTLXDS--SEKDGL-----        354  
  
Qy   491 NPTATIOATAIRIVEPESDSEKLGSKGLDNSSQSGSLPET--PEHSSG-----     538  
    |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::||  
Db   355 ---SEVDTVA---ADPSALEED-----REPPEATGTGAHPGHGGHPPSLANGAA 399  
  
Qy   539 -BEDDWESSADEAESLKLWNSFCNSDDPNLFNKARFQTSGENEKGCRDCKTSPESI VA 597  
    |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::||  
Db   400 QDGTHPESTGESDS-----SPRQGGDHSCGCC-DASCCSFPCYS 438  
  
Qy   598 ISECHITLLSKCVQLLGSQSECPDSVORDVLSSGR---HTHVGRKKVTFLEEVETYYISG 654  
    |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::||  
Db   439 -SSCYS-TSCYSSSCYSASCVP-----SCYNGNRFASTHFSSVDSAKISESTVPSSOD 491  
  
Qy   655 DEDRK-----GW-BEFARDGRCFKRIQETEDA-I-----          683  
    |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::||  
Db   492 DEERNASFZSVPMQSMOSPELDPESTNGAGPWQDELAAPSGVHERSPGLESPVAGPSNR 551  
  
Qy   684 --GYCLTFEHRERMFNRLQCTCFKGLNVLYKC 713  
    |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::||  
Db   552 REGFTTADASKQGFNM---TC-----LI RDC 575
```

```
RESULT 15  
PCT-US03-10870-2173  
; Sequence 2173, Application PC/TUS0310870  
; GENERAL INFORMATION:  
; APPLICANT: Mitokor, Inc.  
; APPLICANT: Buck Institute  
; APPLICANT: Ghosh, Soumitra S.  
; APPLICANT: Fahy, Erin D.  
; APPLICANT: Zhang, Bing  
; APPLICANT: Gibson, Bradford W.  
; APPLICANT: Taylor, Steven W.  
; APPLICANT: Glenn, Gary M.  
; APPLICANT: Warlock, Dale E.  
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION  
; FILE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME  
; FILE REFERENCE: 660088.465PC
```

```

; CURRENT APPLICATION NUMBER: PCT/US03/10870
;
; CURRENT FILING DATE: 2003-04-04
;
; NUMBER OF SEQ ID NOS: 325
;
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 2173
; LENGTH: 1857
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US03-10870-2173

```

[illegible]

Search completed: September 16, 2005, 10:56:21  
Job time : 88.4089 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 16, 2005, 08:56:27 ; Search time 18.1914 Seconds  
(without alignments)  
3771.163 Million cell updates/sec

Title: US-10-650-482-2

Perfect score: 3804

Sequence: 1 MFPGTGGSKRLGLRAGFRF.....RMFNRLQGTCTGKGLNVLKQC 713

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

PIR\_79:\*

1: pir1:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	237	6.2	657	2 S10001	MyD116 protein - m
2	201	5.3	590	2 A56535	gadd34 protein - l
3	159	4.2	757	2 S68142	Probable transcrip
4	150.5	4.0	2282	2 T42717	DNA-binding protei
5	150	3.9	1062	2 T46444	hypothetical prote
6	149	3.9	647	2 S58225	skeletal muscle ab
7	148	3.9	1567	2 T03730	antigen containing
8	145	3.8	2774	2 A43359	microtubule-associ
9	144.5	3.8	3122	2 T17202	DNA-directed DNA p
10	143.5	3.8	574	2 S37762	25K protein - Bab
11	142.5	3.7	1621	2 A82255	hypothetical prote
12	142	3.7	406	2 S38170	SRP40 protein - ye
13	141.5	3.7	833	2 T22139	hypothetical prote
14	139	3.7	846	2 T25234	hypothetical prote
15	139	3.7	1346	2 A57376	probable regulator
16	139	3.7	1560	2 T30282	calcium-binding pr
17	139	3.7	5170	2 T15348	hypothetical prote
18	138.5	3.6	1271	2 A45555	glutamate rich pr
19	138	3.6	1311	2 G86471	unknown protein [i
20	137	3.6	1075	2 B96508	hypothetical prote
21	136.5	3.6	532	2 T06029	hypothetical prote
22	136.5	3.6	1807	2 S03124	vitellogenin A2 pr
23	136	3.6	772	2 A55004	transcription fact
24	136	3.6	1437	2 F96783	unknown protein F2
25	136	3.6	1736	2 T00391	hypothetical prote
26	135.5	3.6	534	2 T39903	serine-rich protei
27	135	3.5	1304	2 T19397	hypothetical prote
28	134.5	3.5	2218	2 B84683	hypothetical prote
29	134	3.5	1046	2 T29776	hypothetical prote

30	134	3.5	1830	2 A37981	microtubule-associ
31	133	3.5	854	2 T17288	hypothetical prote
32	133	3.5	917	2 B85057	hypothetical prote
33	133	3.5	933	2 S41539	fibrinogen-binding
34	133	3.5	1825	2 S13507	microtubule-associ
35	133	3.5	4307	2 T20721	hypothetical prote
36	132.5	3.5	966	2 S25365	CYC8 protein - yea
37	132.5	3.5	1447	2 S50918	DNA helicase TRS1
38	131.5	3.5	1229	2 T46116	hypothetical prote
39	131.5	3.5	1435	2 A37793	erythrocyte-bindin
40	131	3.4	736	2 T00023	transcription fact
41	131	3.4	772	2 S62481	hypothetical prote
42	131	3.4	1048	2 S64758	major merozoite su
43	131	3.4	1772	2 A45532	hypothetical prote
44	131	3.4	2761	2 T21064	hypothetical prote
45	130.5	3.4	1979	2 J00059	mtprd protein - mo
46	130.5	3.4	2364	2 A56577	microtubule-associ
47	130.5	3.4	2416	2 T13825	adenomatous polypo
48	130	3.4	630	2 S29796	hypothetical prote
49	130	3.4	877	2 F90070	Clumping factor B
50	130	3.4	1262	2 T25168	hypothetical prote
51	130	3.4	1805	2 A34736	hypothetical prote
52	129.5	3.4	1108	2 A48508	cyclic-nucleotide
53	129	3.4	727	2 S18193	autoantigen NOR-90
54	129	3.4	727	2 JCS113	ribosomal transcri
55	129	3.4	764	2 S09318	transcription fact
56	129	3.4	764	2 JCS112	ribosomal transcri
57	129	3.4	1664	2 T18216	integrin-like prot
58	129	3.4	2342	2 T13412	hypothetical prote
59	128.5	3.4	184	2 A42749	hypothetical prote
60	128.5	3.4	1810	2 E88481	protein Cl6A3.2 [1
61	128.5	3.4	7160	2 T27935	hypothetical prote
62	128	3.4	756	2 T00367	hypothetical prote
63	128	3.4	989	2 D89852	fibrinogen-binding
64	128	3.4	990	2 I51618	nucleolar phosphop
65	128	3.4	1024	2 E86331	hypothetical prote
66	128	3.4	1087	2 S28282	hypothetical prote
67	128	3.4	1093	2 F85556	hypothetical prote
68	128	3.4	1684	2 J00057	protein B0464.5a [
69	128	3.4	3924	2 S37431	gravin - human
70	127.5	3.4	490	2 G81723	ankyrin 2, neurona
71	127.5	3.4	568	2 I39411	conserved hypothet
72	127.5	3.4	820	2 T45685	AF-9 protein - hum
73	127.5	3.4	878	2 T17245	hypothetical prote
74	127.5	3.4	1463	2 T30290	AAS surface protei
75	127	3.3	589	2 S74668	hypothetical prote
76	127	3.3	781	2 T41551	hypothetical prote
77	127	3.3	1213	2 A58198	serine/proline-ric
78	127	3.3	1317	2 T03748	apoptosis associat
79	127	3.3	1472	2 A84470	hypothetical prote
80	127	3.3	1510	2 T33100	hypothetical prote
81	127	3.3	2327	2 T42630	aggreccan - bovine
82	126.5	3.3	889	2 T11742	egg sperm receptor
83	126	3.3	710	2 S30154	low-temperature-in
84	126	3.3	1300	2 T03166	probable immediate
85	125.5	3.3	1257	2 S28764	neobcan precursor
86	125.5	3.3	2132	1 A55182	aggreccan precursor
87	125	3.3	482	2 A44997	merozoite surface
88	125	3.3	636	2 T51893	related to Che-1 p
89	125	3.3	817	2 S53319	hypothetical prote
90	125	3.3	903	2 T47316	hypothetical prote
91	125	3.3	1804	2 T34518	nestin - golden can
92	125	3.3	1812	2 I49350	breast/ovarian can
93	125	3.3	3375	2 T19821	hypothetical prote
94	125	3.3	5762	2 A41819	proline-rich pep1
95	124.5	3.3	968	2 T45746	hypothetical prote
96	124.5	3.3	1113	2 T47381	hypothetical prote
97	124.5	3.3	1153	2 F84468	hypothetical prote
98	124.5	3.3	1325	2 S62497	probable nucleopor
99	124.5	3.3	1355	2 S40022	spalt protein - fr
100.	124.5	3.3	1658	2 S55101	hypothetical prote



C;Accession: 146444  
 R;Bloecker, H.; Boecher, M.; Brandt, P.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.  
 submitted to the Protein Sequence Database, January 2000  
 A;Reference number: Z23032  
 A;Accession: T46444  
 A;Status: preliminary

A;Molecule type: mRNA  
A;Residues: 1-1062 <AAA>  
A;Cross-references: UNIPROT:Q9NSW2; EMBL:AL137701  
A;Experimental source: adult testis; clone DKFZp434N1427  
C;Genetics:  
A;Note: DKFZp434N1427.1

Query Match 3.9%; Score 150; DB 2; Length 1062;  
Best Local Similarity 20.5%; Pred. No. 0.25;  
Matches 116; Conservative 71; Mismatches 187; Indels 192; Gaps 25;

```
QY 223 LNPYSIDCFPRLEVSQNSDGNSEVVGFTLTTPSSCLREDCPKQPLSAELLIPASMOGC 282
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 18 LNPRLSSIP-----LSSTPCHLSPSSLPFSVAERKP 50

QY 283 PPLSTEGPLPEIHLMKRLKLEFLQANKG-----QDLTPDQDNGVHSLER 327
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 51 PLFNMMASALYHIAQNSPVLQSHGEYFRNFVDSCLQIKIPDRPTSE-----VLLK 104

QY 328 EHSILRMDPKHCRDNPQTQVPAAGDIPGNTQESTTEKIELLTTEVPLALEBESPEGCP 387
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 105 HRFLRERP-----PTVIM-----DLIORTKDAVRELNLQVRMKKILFQEAENG--PG 152

QY 388 SEIPMEKEPGE-----GRISVVDYSYLEGDLPIASARPACS--NKLIDYILGGASDLET 439
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 153 AEAPEREERAPYMRACGLTSLSSHVSVPMSISASSQSSVNSLAD-----ASDNE- 205

QY 440 SSDPEGEWDDEADDGDFDSSLSDSLQDPGRGLHLWNFCV-----DPY 487
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 206 -----EEEEEEEBEGPEA-----REMAAMMOEGEHTVTGSHSIHRLPGSDNLYDDPY 257

QY 488 NPQ-----NFTATIQTA-----ARIVPEEPSDS--EK 512
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 258 QPETPSPLOPPAAPAPTSTSSARRRAYCNRDHAFATIRASIVSRIOQHEQDSALRE 317

QY 513 DLSGKSDLNENSSQSGSLP-----ETPEHSSGDEDDWESS-----ADAEASLKLWNFCN 561
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 318 QLSGKVMRRHQKQLLALESRLRGEREHSARLQRELEAQRAGFAEAKL----- 369

QY 562 SDDPYNPLNFAPQPTSGENKGRDSTKTPESIVASECTHLISCKVQLLGSQSECPD 621
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 370 -----ARRHQAIKEKE--ARAAQAEERKF-----QQHILGQOKGEL-- 403

QY 622 SVQRDLVSGGRHTVVRKVTFLERVTEYIISGDEDRKGPWEERFARDGCRFKRIO--ET 679
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 404 ---HALEAQKRTY-KLKEQIKEL-----QENSTPRKRAEMLLRKQELQOCQA 452

QY 680 BDAIGYCLTFEHRERMFNLQGTCPK 705
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 453 EEEAGLL---RRQRQYFELQCQRYK 474
```

RESULT 6  
S58225  
A;Molecule type: protein - human  
A;Residues: 1-647 <NIE>  
A;Cross-references: UNIPROT:Q969M9; EMBL:X87613  
C;Superfamily: bromodomain homology  
C;Keywords: skeletal muscle  
F;506-561/Domain: bromodomain homology <BRO2>

Query Match 3.9%; Score 149; DB 2; Length 647;  
Best Local Similarity 20.5%; Pred. No. 0.15;  
Matches 123; Conservative 86; Mismatches 213; Indels 178; Gaps 30;

```
QY 56 PETRVSYMTKLISQLLAPLGLLQKVLWSQLFGGMEPTR----- 95
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3 PVTDDSPQKMWLGKATPPSPPLSELKK---GSLLPSPRLVNESEMAVASGHLNSTG 59

QY 96 -WLDFAGVYSALRALKGREKP---AAPTAQ---KSLSSQLQDSDSDPSPVSPDLWLEGIH 148
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 60 VLLEGGVLPMIHGGEIQOTPTNTVAASPAASGAPTLISRL--LEAGPTQFTPL-----ASFT 114

QY 149 WQYSPDPLKLELKAKGALDPAQAQFLLEQOOLMGVELLPSSLOQRRLYSNRELSSPGPL 208
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 115 TVASEPPVKL---VPPPEVESVQATIVM-----MPALPAPSSAPAVSTTESVAVSQPD 165

QY 209 NIQRID-----NFSVVSYLNPVLDQCF-----PRLE----- 235
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 166 NCVPEAVGDPHTVTVMDSSEISMINSIKBEFCRSGVABAPVGSKAPSIDGKEELDLA 225

QY 236 -----VSYQNSDGNSEVVGFTLTTPSSCLREDC--HPQPLSAELLIPASMOGC---P 283
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 226 EKMDIAVSYTGEELDFETVG-----DIIAIEDKVDHPEVLDVAVAEAAALSFCENDD 279

QY 284 PLSTEGPLPEIHLMKRLKLEFLQANKGQDLPTPD-----QDNGYHSL---EE 327
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 280 POSLPG-PWEHPIQ-----QERDKFVPLPAPETVYKQERLDFEETENKGIHELVDIRE 331

QY 328 EHSILRMDPKHCRDNPQTQVPAAGDIPGNTQESTE-----EKI-----ELLTTRVPL 374
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 332 PSABIKVEAPBPPEVIVGASIVAGVVPATSNPEPLRSQDLDELGSTAGAILLEADVAI 391

QY 375 ALHEESP-----SBCGSPSEIPMEKEPGEGRISVVDYSYLEGDLPIASARPACSNKLIDYI 429
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 392 GKGDTEPLTNVKTAEASPESML---SPSHGSPNPIED-----PLEAETQHKFEMSDSL 439

QY 430 LGGASSDLETS-SDPEGEDWDEAEADDGDFDSDSLSD-----SDLEQDPEGHL 477
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 440 KEESGTIFGSIQADAPGED---EEDGVSEAAALEEPKEEDQEGYLSMDNEBPVSE- 494

QY 478 WNSFCSDVPYNPNQFNFTATIOAARIVPEEPSDSE--KDLSG-KSDLNENSSQSGSLPTPE 534
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 495 -----SDGFSIHNATLQSHTLADSISSPASSQLPMDLSTIKKNIEN---GLIRSTAE 545
```

RESULT 7  
T03730  
antigen containing epitope to monoclonal antibody MMS-85/12 - mouse (fragment)  
C;Species: Mus musculus (house mouse)  
C;Date: 24-Mar-1999 #sequence\_revision 24-Mar-1999 #text\_change 09-Jul-2004  
C;Accession: T03730  
R;Fisher, L.W.; Merrelli, A.; Benayahu, D.  
submitted to the EMBL Data Library, July 1997  
A;Description: Clone containing epitope to mouse osteoblast monoclonal antibody MMS-85/1  
A;Reference number: Z15035  
A;Accession: T03730  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: mRNA  
A;Residues: 1-1567 <PIS>  
A;Cross-references: UNIPROT:O35243; EMBL:AF013969; NID:G2384710; PIDN:AAB69856.1; PID:G2

Query Match 3.9%; Score 148; DB 2; Length 1567;  
Best Local Similarity 21.4%; Pred. No. 0.57;  
Matches 105; Conservative 79; Mismatches 173; Indels 134; Gaps 25;

```
QY 238 YQNSDGNSEVVGFTLTTPSSCLREDC-----CHPQPLSAEL---IPASWQGPPLST 287
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1028 HQEPESNLKTT-TKCITGQESQMPSSHTGVLSAVCHVAPCASQEGGLPTKSHSGTWS 1086

QY 288 EGLPE-IHHLRMKRLKLEFLQANKGQDLPTDQDNGYHSLSEEEHSLRLMDPKHCRDNPQTQ 346
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1087 EGSPKMGHVAGARQSFHREGN--LDVTLPPEDNGCGVGNEE-----SPPKG 1131

QY 347 VPAAGDIPGNTQESTTEKIELLTTEVPLALEEE-----SPSEGC-----PSSEIPMEK 394
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1132 I-----GGLELSTG-----LTTEISVSSEEDTSHGVAAAPENPCVGRRRGAELQMEA 1179
```



```

QY 395 -----EPGEGRISVVDVSYLGLDPLISARPACSNKLDIYILGASSDLETSSDPEG 445
DB 1180 LLMRSLNVEKSESRIEETHFESQKEEI-----CCGRKSGSTEALSGCS-----VEADPEE 1231
QY 446 -EDWDREAEDGDPSDSSLDSDLEQDPEGLHLWNSFCSDVPYNPQNTATIQTAAIRVP 504
DB 1232 VEBEKQISQRNRPDYSSSEELDDSPVL-----DSRIETAQRQYS 1274
QY 505 E-EPDSEKOLSGKSDLENSQSGLPTEPHSGEEDDWESSADEASLKLM-----NSF 559
DB 1275 ETEPHDTKEKNSG--DVEEFS---SVTSKTNSTGLREDRDEFSSSEGTGKTEPNEDDGS 1329
QY 560 CNSDDPYNPLNFXA-----PQTSGENEKGRDSTPSEISVAISECHTLTSCVKQL 611
DB 1330 IKSQEDDHPITIKRRGRPRKYPAETAPKSGE---DSKTETD-ITTVQESSPSGKLKV-- 1383
QY 612 LGSQESCPDSVQRDVLSSGRHTHVKKVTFLEEVTEYISGDEDRKGVWEFARDGCR 671
DB 1384 --SQADE-----SNKEIANLEEKs-----TGNDDSEKTKASMLRGRK 1419
QY 672 FQKRIQETEDA 682
DB 1420 PKRSLTSSDDA 1430

RESULT 8
A43359
microtubule-associated protein MAP1A - rat
C:Species: Rattus norvegicus (Norway rat)
C>Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 09-Jul-2004
C:Accession: A43359; S22108
R:Langkopf, A.; Hammarback, J.A.; Muller, R.; Vallee, R.B.; Garner, C.C.
J. Biol. Chem. 267, 16561-16566, 1992
A:Title: Microtubule-associated proteins 1A and LC2. Two proteins encoded in one messenger RNA.
A:Reference number: A43359; MUID:92355629; PMID:1379599
A:Accession: A43359
A:Molecule type: mRNA
A:Residues: 1-2774 <LAN>
A:Cross-references: UNIPROT:P34926; GB:M83196; NID:G205537; PIDN:AAB48069.1; PID:G205538
A:Note: sequence extracted from NCBI backbone (NCBIN:111039, NCBI:P.111040)
R:Cravchik, A.
submitted to the EMBL Data Library, June 1992
A:Reference number: S22108
A:Accession: S22108
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 73-364, 'NRLRS', 370, 'QKN', 374, 'PSPKGL', 381-751, 'RSMMSQMAQR', 764, 'D', 766, 'LR', 'WLKRNMCQPRQSP', 851, 'V', 853, 'NSL', 855, 'LPHRWLRN', 865, 'W', 867, 'HSQLPDGDD', 877, 'Q', 879, 'L'
A:Cross-references: EMBL:X66840
A:Experimental source: strain Sprague Dawley
C:Superfamily: microtubule-associated protein MAP1B
C:Keywords: microtubule binding; phosphoprotein

Query Match 3.8%; Score 145; DB 2; Length 2774;
Best Local Similarity 20.4%; Pred. No. 1.9;
Matches 130; Conservative 78; Mismatches 225; Indels 204; Gaps 29;

QY 109 LKGRKPAAPTAQKSLSLQDSD-----PSVTSPLDWLEEGTH---W-QYSPDPLKL- 158
DB 1126 LRFTDQSLDEADSLSVLSVSPDTTKQKATPRSPCSLKEQPHKDLMPWVSPEDTQSL 1185
QY 159 ----ELKAGSALDPAQAFLLEQLQWGLVLLPSLOSRLYSNRELSSPSGGLNIQRID 214
DB 1186 SPSEFSKETSLSISK-----QLSPESLTLQFGLNLKEERGFPVKAEED 1234
QY 215 NF-----SVVSYLNLNPSYLCFF-----RLVSYQNSDGNSEVWGFTLTTP 255
DB 1235 SCHLAPVSIPEPHRATVSPSTDEPAGTLPGGSPSHSALSVDKHSPEITGPGGHWTs 1294
QY 256 ESSCLRDHCHQPPLSA---ELIPASWQCQPLSTEGLPETHHLRMKRLFLQQAQKQD 312
DB 1295 DSSLTKS-----PESSLSPAMEDLAVEWEGKAP-GKEKEPEL-----KSETROQ--KGQI 1341

```

```

QY 313 LP-----TPQDNGYH-----SLEBEHSLLRMDPKHCRDNPFTQFVPAAGDIPGNTQESTEEK 364
DB 1342 LPEKVAVVEQDLIIHQKDGALDENK-----PGRQDKTPEQ 1378
QY 365 IELLITTEVPLALEEESPSCSPSEIPEMEKEGEGRISVVDVSYLSEGLDPLISARPACSNK 424
DB 1379 -----KGRDLDEKDTAAELDKGPEKPKDLREDQOQORAGP-----PAEKOK 1420
QY 425 LIDYILGASSDLETSSDPEGEDWDDEAEDGDPDSSISLSDSDLEQDPEGLHLWNSFCV 484
DB 1421 ASQ-----RDTDLQOQTATEPRDRAQERRDSE-EKDKSLELRD----- 1458
QY 485 DPYNPNQNTATIQTAAIRV-----PE-----EPSDSEK--DLSGKSDLENSSGSLPET 532
DB 1459 -----RTPEKDRILVQEDRAPEHSIPEPTQTDRAPDRAKGTDDKEQKEBASEEKE- 1508
QY 533 PEHSSGEEDDW-----ESSADEAESIKLWNSFCNSDDPYNPLNFKAPFOTSGENEKGR 586
DB 1509 ---QVLEQKDMALGKGETLDQEARAEQKDETLKED-----KTQOKSGSFVE 1553
QY 587 DSKTSPSESV-----AISECHTLTSCVKQLLGSQSECEPDSVQR-----D 626
DB 1554 DKTTSKETVLDQKSAEKADSVQDQGALEKTRALGLESPAESKARQEKKYKWEQD 1613
QY 627 VLSGGRHTHVKKVTFLEEVTEYISGDEDRKGPWE 663
DB 1614 VVQGWRETSPTRGEP-----VGGQKEPVPAAWE 1640

RESULT 9
T17202
DNA-directed DNA polymerase (EC 2.7.7.7) zeta chain - mouse
C:Species: Mus musculus (house mouse)
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T17202
R:Van Sloun, P.P.H.; Romeijn, R.J.; Eken, J.C.J.
Mutat. Res. 433, 109-116, 1999
A:Title: Molecular cloning, expression and chromosomal localisation of the mouse Rev31 gene.
A:Reference number: Z18720; MUID:99202265; PMID:10102037
A:Accession: T17202
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-3122 <VAN>
A:Cross-references: UNIPROT:Q61493; EMBL:AF083464; NID:G4079830; PID:G4079831; PIDN:AAC09
A:Experimental source: strain 129/Ola; testis
C:Genetics:
A:Map position: 10
C:Keywords: nucleotidyltransferase

Query Match 3.8%; Score 144.5; DB 2; Length 3122;
Best Local Similarity 19.6%; Pred. No. 2.4;
Matches 142; Conservative 95; Mismatches 224; Indels 265; Gaps 37;

QY 70 LIAPLPLGL-----LQKVLINWSQLFG-----GMPTRWLDPAGYSA 105
DB 13 MASPLGLDTCOSPLTQLPVKVPV-VRVFGATPAGQKTCLEHLHGFVLYVYDGY-- 68
QY 106 LRALKGREKPAAPTAQKSLS---SLQDSSDPSTVSP-----LDWLEGIHW 149
DB 69 -----GQGPESVLSQWAFSIDRALNVALGNPSSAQHVKVSLVSGMPPYGTHERHF 122
QY 150 Q-----YSPDLK--LEBKAKGSALD-----PAAQAFLLLEQOLWGLVLL- 186
DB 123 MKIYLYNPAMVKRICELLOSQAIMNKKYQCPHEAHPVLLQLFDYNYLGNLNLAAVKP 182
QY 187 -----PSSLOSRLYSNRELSSPSG-----PLNQRIDNFSVWYLLNPSY 227
DB 183 RKARRKGNASHATGLFKHQLSGNSPAGTLFRWEEDBIPSLLEGEVPLETCELEVDVA 242
QY 228 LDCFPRLVSYQNSDGNSEVWGFTLTTPESSCLREDH-----CHQPLSAELIPASWQ 281
DB 243 ADILNRLDIEAQ-IGGNP---GLQAIWEDEKQRRNRNNESSQISQSPESQDCRFVPAT--- 295

```

```
Qy 282 CPPLSTEGLEPIHILMRKRL-EFLQANKGQDLP-TPDQDNGYHSLEEE---HSLLRMDP 336
Db 296 -----ESKQFQKRLQEVLKQNDFSVTLGSDVDYNGSQEFSAEILHSEI-LSP 344
Qy 337 KHCRRNPQFVPAAGDI---PQNTQESTEEKIELLTTEVPALAEBSPEGCPSEIPME 393
Db 345 EMLPCSPANMIEVHKDITDLSKNTKHKVEE--ALINBEAILNLNLSQT----- 391
Qy 394 KEGGGRISVVDYSYLEGDLPIARPACSNKLIIDVILGASD-----LE 438
Db 392 FOPLTORLS-----ETPVFMGSSPDESILV-HLLAGLES DGYGKKNMPLPCHSFG 441
Qy 439 TSSDPGEGDWDEARDGDFDSSLS-----DSLDQDDPEGLHLMNSFCSDVPYNQN 491
Db 442 ESQPNQSD-DEENEPQIEKEEMELSVVMQSRWSDIE----- 478
Qy 492 FTATIQTAARIVPEEPSSEKDLGSKDLENSSGSLPETPEHSSGGEEDWESADAE 551
Db 479 -----EHCAKRSILCRNAHRSSYTEDD-SSSEBEME 508
Qy 552 SLKLM--NSFC-----NSDDPNPLNFKAPQTSGENEKGCR-----DS 588
Db 509 ----WTDNSLLFANLIPQLGTADENSDNPLNNSRAHSSVATSKLSVRPSIFHKDA 564
Qy 589 KT--PSESIVAISECHTLLCKVQLLG----SQSECPDSVQRDVLGGRHVRKRVKVT 642
Db 565 ATLEPPSSAKITFECKHTSALSSHVNLKDGTLTDLQSPNSTEK-----GRD-----NSVT 614
Qy 643 FLEEV 648
Db 615 FTKEST 620
```

## RESULT 10

```
S37762
225K protein - Babesia bovis (fragment)
C:Species: Babesia bovis
C:Date: 19-May-1994 #sequence_revision 01-Dec-1995 #text_change 09-Jul-2004
C:Accession: S37762
R:Jasmer, D.P.; Reduker, D.W.; Perryman, L.E.; McGuire, T.C.
Mol. Biochem. Parasitol. 52, 263-270, 1992
A:Title: A Babesia bovis 225-kilodalton protein located on the cytoplasmic side of the
A:Reference number: S37762; MUID:92319010; PMID:1377786
A:Accession: S37762
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-574 <JAS>
A:Cross-references: UNIPROT:Q00708; EMBL:M80466; NID:g155856; PIDN:AAA27791.1; PID:g1558
```

```
Query Match 3.8%; Score 143.5; DB 2; Length 574;
Best Local Similarity 20.3%; Pred. No. 0.27;
Matches 109; Conservative 78; Mismatches 194; Indels 157; Gaps 25;

Qy 57 ETRVSYWTKLSQLLAPLGLLQKVLINSQLFGGMPTTRWLDFAGVYSALRALKGREKPA 116
Db 136 KVRIHRVNYLPGLLQQLPDYLIKALKCNIVEGISPDVMLEI---EYLLKVNKNKPM 191
Qy 117 APTAKSLSSQLDSSD-----PSVTSPLDWLEEGHWHQYSPDLKLEKAKGSALDP 169
Db 192 SPEIIRDTNDEMMDTKPAQVPTTPSP-----SITDE 226
Qy 170 AAQAPLLQQLGWVLELLPSSQLSRLYSNRELGSPGPNLQRIQDNFVSVYLLNPSYLD 229
Db 227 VIQ---LPEYVDGVPVFFSHIDDEI-----IQLED-----YESPVLVP 262
Qy 230 CFPRLVSYQNSDGNSEVVGFTLTTPESSCLRE-----DHCHPOPLSAELIPA 277
Db 263 IETEIGSESSSSSDSEGEITLPEDSGVDPDDVDVQLPEYVDSPIVLPIETEI--- 319
Qy 278 SWQCFFPLSTGLPEIHILMRKLEFLQANKGQDLPDQDNGYHSLEEHSLLRMDPK 337
Db 320 --EGSESESSG-----SDSEGEITLIP-BDSGVE---EPDDDVQL-PE 357
```

```
Qy 338 HCRDNPQFVPAAGDIPGNTQESTEEKIELLTTE-VPLALEEBSPEGCPSE----- 389
Db 358 YV-DGPVPF-PSHD-----DEIIQLPDYEDSDPIVLPIETIEIGSESSSSSDSE 406
Qy 390 -----IPME---KEPGGRISVVDYSYLEGDLPIARPACSNKLI-----DY-----ILGAS 434
Db 407 GEEITLPEDSGVBEPPDDVQLPEYV---DGPVVPFSDHIDDEIIQLPDYEDSDPIVLPIE 462
Qy 435 SDLETSDDPGEWDDEAEADGDFDSSLS--DSLDQDDPEGL-----HLMNSFCSV 484
Db 463 TEIGSESSSSSGSDSEGEITLPEDSGVEEPDDVDVQLPEYVDPVVPVFFSHIDDEIIQL 522
Qy 485 DPYNPQNFTATIQTAARIVPEEPSSEKDLGSKDLENSSGSLPETPEHSSGGEED 542
Db 523 PDYEDSDPIVLPIET-----EIGSESSSSSSSDSE-----GEEITLPEDSGVVEED 568

RESULT 11
A82255
hypothetical protein VC0998 [imported] - Vibrio cholerae (strain N16961 serogroup O1)
C:Species: Vibrio cholerae
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 09-Jul-2004
C:Accession: A82255
R:Heidelbergl, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.J.;
chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, P.
L. R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: A82035; MUID:20406833; PMID:10952301
A:Accession: A82255
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1621 <HEI>
A:Cross-references: UNIPROT:Q9KTA5; GB:AE004181; GB:AE003852; NID:g9655454; PIDN:AAF9415
A:Experimental source: serogroup O1; strain N16961; biotype El Tor
C:Genetics:
A:Gene: VC0998
A:Map position: 1
```

```
Query Match 3.7%; Score 142.5; DB 2; Length 1621;
Best Local Similarity 20.5%; Pred. No. 1.3;
Matches 155; Conservative 100; Mismatches 278; Indels 223; Gaps 35;

Qy 33 SSXFPFP-----LGPENSGNPTL-----LSSAQ---PETRVSYWTKLSQLLAPLPG 77
Db 779 ANEGCTPQDEWDVDEDDSSFTLEGNAELSAEDDLPE-QTTATNETADELLADLAA- 836
Qy 78 LQKVLINSQLFGGMPTTRWLDFAGVYSALRALKGREKPAAPTAKSLSSQLDSSDPSVT 137
Db 837 -----QPQSNVTVDTSDDALAPDGLSQSVE 860
Qy 138 SPLD-----WLREGIHWHQYSPDPLKLELKAQSALDPAA--- 171
Db 861 EPLTLNDLLEPERNDEPQLAEVTPSSAFDEQQVETEIEPESEPLAAEASNDLSLTALNE 920
Qy 172 ---QAFLEQQLGWVLELLPSSQLSRLYSNRELGSPGPNLQRIQDNFVSVYLLNPSYL 228
Db 921 LDLPYTEEDVLADVQLEPAA-ESEVPDLVNEPVEEAFTELDLDELDPYTEEDAL- 978
Qy 229 DCFPRLEVSYQNSDGNSEVVGFTLTTPESSCLRDHCHPOPLSAELIPASWQCQCPPLSTE 288
Db 979 -----ADAQLEPVAESEVEPELDLASE-----PABEEAFTELNKL 1013
Qy 289 GLPEIHH---LRMKRLEFLQANKGQDLP---PTPDQDNGYHSLEEHSLLRMD- PKHCRD 341
Db 1014 DLPEYTEEDALADAQLESATESVESELELVSPAAREAFTELDE-----LDLPYTEE 1067
Qy 342 N---PTQFVPAAGDIPGNTQESTEEKIELL-----TTEVPLA--- 375
Db 1068 DALADSLQLEPAA-----ESEVEPELELVSEPVTEEAFTELDLDELDPYTEEDALADAQ 1120
Qy 376 LEESPSGCPSPSEIPMEKEPGEGRISV-----VDY-SYLEGDLPI SAR--PACSNKLIDY 428
```

```
Db 1121 LEPAVESEVP--BLELASEPABEASTELNELDLPEYTEEDALADAQLEPAASEV--- 1175
Qy 429 ILGGASDLTSSDPGE-----DWEEEDDGFDSDSLSDLEQ-----DPE 473
Db 1176 -----ESELEASLEEEEPTELNELDLPEYTEED-----ALADAQLEPAVESEVEPE 1224
Qy 474 GLHL-----WNSFCSVDYPNPQNFATIQTA-ARIVPEPSPDSEKDLGSKDLSNSQS 526
Db 1225 -LELATPABEAPPELDELDPYTEEDALADAQLEPAEAEVEPELELASDLEEKETP 1283
Qy 527 GSLPE--TPEHSSGE-----EDDWESSADEASLKLWNSFCSNDDPYNPLNPKA 573
Db 1284 TELDELDPYTEEDALADAQLEPAEAEVEPELELASVLEEDSFTELDELDPYTEK 1343
Qy 574 PFQTSNGENKGRDKTP-----SESIIVASECHTLLSKVQLLGSOESEPDSVQVRDVL 628
Db 1344 DALADAQLEPAEAEVEPELGDTEETLAQETESDALVADE--DLLASVES-AVDEVQPELL 1401
Qy 629 SGRHRTYKRA--KVTFLVEEYTYISGDEDRKGPW 662
Db 1402 GATQDVPTQSLANKAFDEEALHDWLSNDPDKGKPF 1437

RESULT 12
SRP40 protein - yeast (Saccharomyces cerevisiae)
N;Alternate names: protein YKR092c; protein YKR412a
C;Species: Saccharomyces cerevisiae
C;Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 09-Jul-2004
A;Accession: S38170; S40645; S37702
R;Baladron, V.; Ballesta, J.P.G.; Bou, G.; del Rey, F.; Esteban, P.F.; Garcia-Cantalejo,
submitted to the Protein Sequence Database, March 1994
A;Reference number: S38158
A;Accession: S38170
A;Molecule type: DNA
A;Residues: 1-406 <BAL>
A;Cross-references: UNIPROT:P32583; EMBL:Z28317; NID:g486580; PID:g486581; MIPS:YKR092c
A;Experimental source: strain S288C
R;Bou, G.; Esteban, P.F.; Baladron, V.; Gonzalez, G.A.; Cantalejo, J.G.; Remacha, M.; J.
Yeast 9, 1349-1354, 1993
A;Title: The complete sequence of a 15 820 bp segment of Saccharomyces cerevisiae chromo
A;Reference number: S40644; MUID:94205265; PMID:8154186
A;Accession: S40645
A;Molecule type: DNA
A;Residues: 1-406 <BOU>
A;Cross-references: EMBL:X73541; NID:g450550; PIDN:CAA51946.1; PID:g450552
R;Lalo, D.; Carles, C.; Sentenac, A.; Thuriaux, P.
submitted to the EMBL Data Library, May 1993
A;Description: Interactions between three common subunits of yeast RNA polymerases I and
A;Reference number: S37702
A;Accession: S37702
A;Molecule type: DNA
A;Residues: 1-399, 'N', 401-406 <LAL>
A;Cross-references: EMBL:L11275; NID:g295670; PID:g295671
C;Genetics:
A;Gene: SGD:SRP40
A;Cross-references: SGD:S0001800; MIPS:YKR092c
A;Map position: 11R

Query Match 3.7%; Score 142; DB 2; Length 406;
Best Local Similarity 23.4%; Pred. No. 0.21;
Matches 71; Conservative 27; Mismatches 121; Indels 84; Gaps 10;

Qy 370 TEVPLALEEPPSPGCPSSIEIPMEKPEGRISVVVYSYLEGLDLPISARPACSNKLLDIYI 429
Db 146 TEPSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS 191
Qy 430 LGGASSDLETSSDPGEEDWEEAEDDGFDSDSLSDLEQDPPEGLHLWNSFCSVDYPNP 489
Db 192 ESDSQSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS 236
```

## RESULT 13

```
T22139
hypothetical protein F43G6.9 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
A;Accession: T22139; T22174
R;Swinburne, J.
submitted to the EMBL Data Library, July 1995
A;Reference number: Z19522
A;Accession: T22139
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-833 <WIL>
A;Cross-references: UNIPROT:Q20374; EMBL:Z50070; NID:e1519046; PIDN:CAA90402.1; GSPDB:GN
A;Experimental source: clone F43G6
R;Matthews, L.
submitted to the EMBL Data Library, November 1996
A;Reference number: Z19526
A;Accession: T22174
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-833 <W12>
A;Cross-references: EMBL:Z83108; PIDN:CAB05512.1; GSPDB:GN00020; CESP:F43G6.9
A;Experimental source: clone F44E5
C;Genetics:
A;Gene: CESP:F43G6.9
A;Map position: 2
A;Introns: 21/1; 68/3; 122/3; 186/3; 381/3; 474/3; 641/2; 752/3

Query Match 3.7%; Score 141.5; DB 2; Length 833;
Best Local Similarity 18.1%; Pred. No. 0.6;
Matches 139; Conservative 106; Mismatches 251; Indels 273; Gaps 36;

Qy 6 GGSKRKRLGPR-----AGFRFWPPFPFRRSQA-----GSSKPTPLGPNNSGN--- 47
Db 50 GGGLDNIGENAELENTATQAKLRFDPPVWQKSSSDHVAAPSASEIPFP--FNFGNGDA 107
Qy 48 ----PTLLSQAQETRVSVYWKLLSOLLAPLPGLLQKVLWSQLFG----GMPTRWLDF 99
Db 108 SDSFKSFEAESPFLLKKSIGN-----CTGDYAINWNTGNTFGISSVPAAPTLLDLP 157
Qy 100 AGV-----YSALRALKGRKPAAPTAQKSLSS---LQLDSSDPSVTSPLDWLEGI---- 147
Db 158 GALLPTFTTQATKEVKSQIIPSPSALTLEDCERMQMGKGNKPSLDVDAFKQLQLGTFVQ 217
Qy 148 --HWQYSPDPLKLELAKGKSGALDPAQAFLLEQLMGVELLPSSLQSRLYSNRELSSPS 205
Db 218 PQSAQSKKPLEARIAPGTPTATSSQA-----LPTLPT 251
Qy 206 GPLNIQRIDNFSVYLLNPSYLDPCFPLEV--SYQNSDGNSEVGFQTLTPSSCLURE 263
Db 252 AALSLEEL-----LQIMKEAQILKGRQV-----PSDWRED 283
Qy 264 H--CHQOP-----LSAELIPASWQCPLPSTGLPEI-----HHLRMKRL----- 301
Db 284 NKFSPHPGPNQNVQPRMDPSLSPGMGHGMPSMGTSMPHGMPPQNMQLPRLPLNP 343
```

Qy	302	EFLQ-----QANKGQDLPTP----	DQNGYHSLEEHSLRMD--PKHCRDNP	343
		:   :	:   :	
Db	344	QELPLIPVWFNAIINNIQLPMGVPPPPFLFQLLNHYRNLPQLVHAMIMOSSIPPNTIRQG	403	
		:   :	:   :	
Qy	344	TGFV-PAAGDIGPNTQEETEEKIELLTTVPALAEESPSEGCPSSSEIPMEKEPGGRIS	402	
		:   :	:   :	
Db	404	PQSHPSGHPSFGNVRKHSGMPSTRTIYDLAID-	S 439	
		:   :	:   :	
Qy	403	VVDY-SYLEGDLPISARPACSNKLIIDYILGGASDLETSSDPREGD-----WDDEEAEDDG	456	
		:   :	:   :	
Db	440	FAGYMSYKERWLR-----IQFI-----QCKGSGDPQVDDVVYVWRDKQIANG	484	
		:   :	:   :	
Qy	457	FUSDSSLSDSLDEQDPGLHLWNPFCSVDPNYPNQNFATITQTAAIRIVPEEPSDEXKLSG	516	
		:   :	:   :	
Db	485	WTAETKLEATEK-----KEKSSESQNDY---S 509		
		:   :	:   :	
Qy	517	KSDLNSSOSGSGLPETPHSGSED-----DWESSADEAESLKLN-----S 558		
		:   :	:   :	
Db	510	---LERISM-NYREMQERARERDKERQREQEIRDGEDKKLRQTLSDKFATSLGLPS	565	
		:   :	:   :	
Qy	559	FCNSDDPYNPLNFKAPFTSGENEGKGRDSKTPSESIVAJISECHTLSCVKVLGSG----	614	
		:   :	:   :	
Db	566	KSSTNPRHVLOMKAVESVDNQTKLSDSE--RKIAVAKRLTML---LRLEGALNL	619	
		:   :	:   :	
Qy	615	-----QSECEPDVSQ-RDVLSGGRHTHVKKVTFFLEEVTEYYISGDE	656	
		:   :	:   :	
Db	620	MEVDELRRSLPEKSQFKDLSDEKDQVEKRVTVIINE-----LMGDD	663	
		:   :	:   :	

RESULT 14  
T25234  
hypothetical protein T24D1.2 - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C;Accession: T25234  
submitted to the EMBL Data Library, October 1996  
A;Reference number: Z20001  
A;Accession: T25234  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: DNA  
A;Residues: 1-846 <SWIL>  
A;Cross-references: UNIPROT:O02331; EMBL:Z81131; PIDN:CAB03423.1; GSPDB:GN0001  
A;Experimental source: clone T24D1  
C;Genetics:  
A;Gene: CESP.T24D1.2  
A;Map position: 1  
A;Introns: 76/2; 125/2; 208/2; 258/2; 338/2; 369/3; 399/2; 517/1; 539/2; 749/2;

Qy 549 E-AESLKLNWSCNDDPPYNLFKAPQTSGE-----NEKGRDSDTKTPSSSIVA 597  
 Db 280 SEISDSYMSYSDGFGTSMPIRTCTPDGTWGDCTMCCEVPTEPGQC-----NRCQIIIG 335  
 Qy 598 ISEC-----HTLLSKCVQLLGSESECPDSVQRDLVSGG 631  
 Db 336 CSACIVHHHVALSPSCPLCRRRWSRQPDVSNMNVLGIG 374  
  
 RESULT 15  
 A57376  
 Probable regulatory protein 322 - rat  
 C;Species: Rattus norvegicus (Norway rat)  
 C;Date: 08-Feb-1996 #sequence\_revision 08-Feb-1996 #text\_change 08-Feb-1996  
 C;Accession: A57376  
 Mol. Cell. Biol. 15, 2754-2762, 1995  
 R.;Lin, X.; Nelson, P. J.; Frankfort, B.; Tomblar, E.; Johnson, R.; Gelman, I. H.  
 A;Title: Isolation and characterization of a novel mitogenic regulatory gene, 322, which  
 A;Reference number: A57376; MUID:95257957; PMID:7739556  
 A;Accession: A57376  
 A;Status: preliminary; not compared with conceptual translation  
 A;Molecule type: mRNA  
 A;Residues: 1-1346 <LIN>  
 A;Cross-references: GB:U23146  
 C;Genetics:  
 A;Gene: 322  
  
 Query Match 3.7%; Score 139; DB 2; Length 1346;  
 Best Local Similarity 19.6%; Pred. No. 1.7;  
 Matches 145; Conservative 112; Mismatches 296; Indels 188; Gaps 34;  
  
 Qy 4 GTGGSRRRLG--PRAGFRFWPPFFRRRQAASSKPTPLGPNSGNPNTLLSSAQPETRV 60  
 Db 306 GGGQSQRSGQQRRSRTDAVPASTQEQDQAQSSSPPEAGSPSEGE-----GV 354  
 Qy 61 SYTKLLSQLAP----LPGLQKVLIMSQLFGMPPTRW-----LDFAGVYSAL 106  
 Db 355 STW-ESFKRLVTPRKSKLBEKEAGRTLIVVAGCPLNRVKNLGFPLRNSPDGGR 413  
 Qy 107 RALKGKEKPAAPTAQKLSLSLQDSSDPVTS--PLDWLEEGIHWOYSPDLKLELKAG 164  
 Db 414 KGQWGRQEA--IVEDS-GPVEINEPDVPVAVPLS-----EYDAVE-REKMEAQG 461  
 Qy 165 SALDPAAQAPLLQQIWGVELLPSSLOSLRYNSRELSSPGPLNLTQRDNFSVSYL-- 222  
 Db 462 NABLPSG-----WGCVV-----SEELSKTLVHTVSVAVDGTRAVTSVEE 501  
 Qy 223 LNP5YLDGF-----PRLEVSQNSDGNSEVVGFOTLTPESSCLREDHCHPQ 268  
 Db 502 RSPSWLSASVTEPLEHTAGEMPVEEVEKDIABETPVLQTL--PEGDAHDDMWTSE 560  
 Qy 269 -PLSAELIPASWGCPCPLSTEGLPETIHLRMKRLFLOQANKGQDLF-----TPDQ--D 319  
 Db 561 VDFTSNAVTA-TETSEALRTEEVTEASGAE-ETDMVSAVSQLTSDSPDTTEATPQVEVE 618  
 Qy 320 NGVHSLEEE-----HSLRMDPKCRN-----PTQFVPAAGDIPGNTQESTEEKIELLTTE 371  
 Db 619 GGVLDTEEBERQTALQIADVADKKSESQVPATQTQRTGSKALEKVEEVEDSEVLASE 678  
 Qy 372 VPLALBEESPCEGSPSETPMEKEPGEGRISVVVDSYLEGDLPI SARPACS-----NKLID 427  
 Db 679 KEKDVMPKGPVQAGAHLAQGETQATPESLEVEPTADVHDVA--TCQVILKQQLME 736  
 Qy 428 YILGGASDLSTSDPEG-----EDWDEEAADDGFDSDSSLSDSDLBQDPPEGLHWN 480  
 Db 737 QAVAPESSETLTDSETNGSPADSDTADGTQOQDETIDSDSKATAVRQS----- 787  
 Qy 481 FCSVDYPNQNFATTQTAAIRIVPEEPS-----DSEKOLSGKSDLENSQ---SGS 528  
 Db 788 -----QVTEEAATAQGEFPSTLPNNVPAQEEHGEPEGRDVLEPTQOQELAA 835  
 Qy 529 LPETPEHSGEED--DW---ESSADAESIKLWNSCFNDDPPYNLFKAPFOTSGENK 583

Fr1 Sep 16 12:15:26 2005

[illegible]

Search completed: September 16, 2005, 10:46:36  
Job time : 25.1914 secs

**THIS PAGE BLANK (USPTO)**

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	3795	99.8	713	2	O6P156	O6P156	homo sapien
2	3794	99.7	713	2	O96SN1	O96SN1	homo sapien
3	2419.5	63.6	697	2	O8BFW3	O8BFW3	m mus muscu
4	2177	57.2	407	2	O658M2	O658M2	homo sapien
5	1285	33.8	408	2	O6P8G0	O6P8G0	mus musculu
6	1241	32.6	376	2	O8C390	O8C390	mus musculu
7	267	7.0	158	2	O6VZB6	O6VZB6	canarypox v
8	265.5	7.0	674	2	O6IA96	O6IA96	homo sapien
9	260.5	6.8	674	2	O75807	O75807	homo sapien
10	259.5	6.8	674	2	O9NV06	O9NV06	homo sapien
11	237	6.2	657	1	M16 MOUSE	M16 MOUSE	mus musculu
12	217	5.7	578	2	O7TQC2	O7TQC2	rattus norv
13	214	5.6	578	2	O6IN02	O6IN02	rattus norv
14	201	5.3	590	2	O60465	O60465	cricetulus
15	174	4.6	1444	2	O9VFN2	O9VFN2	drosophila
16	174	4.6	1514	2	O8SV55	O8SV55	drosophila
17	170	4.5	1095	2	O9HCG6	O9HCG6	homo sapien
18	169.5	4.5	353	2	O6BQ12	O6BQ12	debaryomyce
19	169	4.4	1596	2	O9Z1P7	O9Z1P7	rattus norv
20	169	4.4	10820	2	O7SDK2	O7SDK2	neurospora
21	167	4.4	1596	2	O62766	O62766	rattus norv
22	162	4.3	4915	2	O6CJB6	O6CJB6	klyveromyce
23	160.5	4.2	1585	2	O9HCC7	O9HCC7	homo sapien
24	160.5	4.2	1614	2	O7QHN89	O7QHN89	homo sapien
25	159	4.2	757	2	O15355	O15355	homo sapien
26	158.5	4.2	5303	2	O9V628	O9V628	drosophila
27	158	4.2	2768	2	O9VC00	O9VC00	drosophila
28	157	4.1	2766	2	O9QZR8	O9QZR8	rattus norv
29	156.5	4.1	919	2	O6ZPY9	O6ZPY9	mus musculu
30	156.5	4.1	1476	2	O9UPF53	O9UPF53	homo sapien
31	156	4.1	746	2	O9B582	O9B582	macaca fasc



```

RESULT 1
Q6P156 PRELIMINARY; PRT; 713 AA.
AC Q6P156;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Protein phosphatase 1, regulatory subunit 15B.
GN Name=PP1R15B;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Eye;
RX MEDLINE=22389257; PubMed=12477932; DOI=10.1073/pnas.2426038999;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins P.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.P., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh P.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.B.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.H.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Eye;
RC Director MGC Project;
RL Submitted (JAN-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC065280; AAG65280.1; -.
SQ SEQUENCE 713 AA; 79125 MW; 76C2DA38F9E76A85 CRC64;

Query Match 99.8%; Score 3795; DB 2; Length 713;
Best Local Similarity 99.7%; Pred. No. 1.3e-201;
Matches 711; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MEGTGGSRKRLGPRAGFRFPFPPRPSQAGSSKFPPTPLGPNNGNTLLSSAQPTRV 60
Db 1 MEGTGGSRKRLGPRAGFRFPFPPRPSQAGSSKFPPTPLGPNNGNTLLSSAQPTRV 60

Qy 61 SYWTKLLSQLAPLGLLQKVLWSQLFGGMPFTRWLDFAGYSAIRALKRGKRPAAFTA 120
Db 61 SYWTKLLSQLAPLGLLQKVLWSQLFGGMPFTRWLDFAGYSAIRALKRGKRPAAFTA 120

Qy 121 QKSLSSQLDSDSPSVTFLDWLEBGIHWQVSPDPLKLELKAGSALDPAQAFLLEQOL 180
Db 121 QKSLSSQLDSDSPSVTFLDWLEBGIHWQVSPDPLKLELKAGSALDPAQAFLLEQOL 180

Qy 181 WGVLLPSSQLRSLYNRELSSPGPLNIQRIQINDFVSVYLLNPSYLDLCPRLVSYQN 240
Db 181 WGVLLPSSQLRSLYNRELSSPGPLNIQRIQINDFVSVYLLNPSYLDLCPRLVSYQN 240

Qy 241 SDGSENVGVFQTLTPESSCLREDHCHPQLSAELIPASWQGCPLSTGLPEIHLRMKR 300
Db 241 SDGSENVGVFQTLTPESSCLREDHCHPQLSAELIPASWQGCPLSTGLPEIHLRMKR 300

Qy 301 LEFLOQANKGQDLPTPDQDNGYHSLSEEHSLLRMDPKCRDNPQTFVPAAGDIPGNTQES 360
Db 301 LEFLOQANKGQDLPTPDQDNGYHSLSEEHSLLRMDPKCRDNPQTFVPAAGDIPGNTQES 360

361 TEEKIELLTTEVPLALREESPSSEGCPSSEIPMEKEPGEGRISVVVDYSYLEGDLPI SARPA 420
361 TEEKIELLTTEVPLALREESPSSEGCPSSEIPMEKEPGEGRISVVVDYSYLEGDLPI SARPA 420

421 CSNKLIIDYILGGASSDLETSSDPGEWDDEAEADGDFDSSLSLSDLEQDPGLHLWNS 480
421 CSNKLIIDYILGGASSDLETSSDPGEWDDEAEADGDFDSSLSLSDLEQDPGLHLWNS 480

481 FCSVDYPNPONFTATIOAARIIVPEEPSDEKDLGSKDLENSQSGLSPETPHSSGEE 540
481 FCSVDYPNPONFTATIOAARIIVPEEPSDEKDLGSKDLENSQSGLSPETPHSSGEE 540

541 DDWESSADEAESLKLWNSFCNSDDPYNPLNFKAPFQTSGENEKGRCDKTPSPSEIVAI 600
541 DDWESSADEAESLKLWNSFCNSDDPYNPLNFKAPFQTSGENEKGRCDKTPSPSEIVAI 600

601 CHTLLSKVQLLGSQSECPDPSVQRDVLSGRHTRHVRKKVTFLEETVTEYISGDEDRKG 660
601 CHTLLSKVQLLGSQSECPDPSVQRDVLSGRHTRHVRKKVTFLEETVTEYISGDEDRKG 660

661 PWEEFARDGCRPKRIQIETSDAIGYCLTFEHRERMFNRLOQTCFGLNLVKQC 713
661 PWEEFARDGCRPKRIQIETSDAIGYCLTFEHRERMFNRLOQTCFGLNLVKQC 713

RESULT 2
Q96SN1 PRELIMINARY; PRT; 713 AA.
AC Q96SN1;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical protein FLJ14744.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=14702039; DOI=10.1038/ng1285;
RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,
RA Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,
RA Sekine M., Obayashi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,
RA Yamamoto K., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahara K.,
RA Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M., Shiratori A.,
RA Sudo H., Hosiiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E., Omura Y.,
RA Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M., Yamazaki M.,
RA Ninomiya K., Ishibashi T., Yamashita H., Murakawa K., Fujimori K.,
RA Tanai H., Kimata M., Watanabe M., Hiraoaka S., Chiba Y., Ishida S.,
RA Ono Y., Takiguchi S., Watanabe S., Yosida M., Hotuta T., Kusano J.,
RA Kanehori K., Takahashi-Fujii A., Hara H., Tanase T., Nomura Y.,
RA Togiya S., Komai F., Hara K., Takeuchi K., Arita M., Imose N.,
RA Yushashino K., Yuuki H., Oshima A., Sasaki N., Aotsuka S.,
RA Yoshihawa Y., Matsumura N., Ichihara T., Shiohata N., Sano S.,
RA Moriya S., Momiyama H., Satoh N., Takami S., Terashima Y., Suzuki O.,
RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,
RA Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,
RA Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,
RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,
RA Ono T., Yamada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y.,
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,
RA Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,
RA Togashi T., Oyama M., Hata H., Watanabe M., Komatsu T.,
RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,
RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuho Y., Yamashita R.,
RA Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.;
RT "Complete sequencing and characterization of 21,243 full-length human
RT cDNAs.";
RT Nat. Genet. 36:40-45 (2004).
DR EMBL; AK027650; BAB55266.1; -.

```

```
DR Genew; HGNC:14951; PPPIR15B.
SQ SEQUENCE 713 AA; 79125 MW; 26C2D06144AAD25E CRC64;
Query Match 99.78; Score 3794; DB 2; Length 713;
Best Local Similarity 99.78; Pred. No. 1.5e-201;
Matches 711; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MEGTGGSRKRLGPRAGFRWPPPPRRSQAGSKFPTPLGPENSGNPTLLSSAQPTRY 60
Db 1 MEGTGGSRKRLGPRAGFRWPPPPRRSQAGSKFPTPLGPENSGNPTLLSSAQPTRY 60
Qy 61 SYWTKLSQLLAPLGLLQVLIWSQLFGGFFRMLDPAGVYSALBAIKGREKPAAPTA 120
Db 61 SYWTKLSQLLAPLGLLQVLIWSQLFGGFFRMLDPAGVYSALBAIKGREKPAAPTA 120
Qy 121 QKSLSSQLQSSDPSVTSPLDWEIEGHWOVSPDPLKLELKAKGSALDPAQAFLLEQQ 180
Db 121 QKSLSSQLQSSDPSVTSPLDWEIEGHWOVSPDPLKLELKAKGSALDPAQAFLLEQQ 180
Qy 181 WGVLELPSLSQRLYNSRELSSPSGPLNQRIDNFSVSYLLNPSYLDGFPRLVSYQN 240
Db 181 WGVLELPSLSQRLYNSRELSSPSGPLNQRIDNFSVSYLLNPSYLDGFPRLVSYQN 240
Qy 241 SDGNSEVVGQTLTPSSCLREDHCHPOPLSAELIPASWOGCPPLSTEGLPETHLMKR 300
Db 241 SDGNSEVVGQTLTPSSCLREDHCHPOPLSAELIPASWOGCPPLSTEGLPETHLMKR 300
Qy 301 LEFLQQAQKQDGLPTPDQDNGYHSLREHSLRMDPKHCRDNPQFVPAAGDIPGNTQES 360
Db 301 LEFLQQAQKQDGLPTPDQDNGYHSLREHSLRMDPKHCRDNPQFVPAAGDIPGNTQES 360
Qy 361 TEEKIELTTTEVPLALEEESPSSECPSSSEIPMEKEPGEGRISVVYDYSYLEGDLPI 420
Db 361 TEEKIELTTTEVPLALEEESPSSECPSSSEIPMEKEPGEGRISVVYDYSYLEGDLPI 420
Qy 421 CSNKLIDYILGGASSDLETSSDPGEWDDEAEADGDFDSSLSLSDLEODPEGLHLWNS 480
Db 421 CSNKLIDYILGGASSDLETSSDPGEWDDEAEADGDFDSSLSLSDLEODPEGLHLWNS 480
Qy 481 FCSVDYPNPQNTFATIQTAARIVPEEPSDEKDLGSKDLENSSQSGSLPETHSHSGEE 540
Db 481 FCSVDYPNPQNTFATIQTAARIVPEEPSDEKDLGSKDLENSSQSGSLPETHSHSGEE 540
Qy 541 DDMESSADEAESLKLWNSFCNSDDPYNPLNFKAPFQTSGENEKCRCDSKTPSEIVAI 600
Db 541 DDMESSADEAESLKLWNSFCNSDDPYNPLNFKAPFQTSGENEKCRCDSKTPSEIVAI 600
Qy 601 CHTLLSKVQLLQSGSECPDSVQDVLSCGRHTHVRKKVTFLEETVEYISGDEDRKG 660
Db 601 CHTLLSKVQLLQSGSECPDSVQDVLSCGRHTHVRKKVTFLEETVEYISGDEDRKG 660
Qy 661 PWEEFARDGCRFKRIQETDAIGYCLTFEHRERMFNRLOQTCFGLNLVKQC 713
Db 661 PWEEFARDGCRFKRIQETDAIGYCLTFEHRERMFNRLOQTCFGLNLVKQC 713
RESULT 3
Q8BFW3 PRELIMINARY; PRT; 697 AA.
AC Q8BFW3;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Mus musculus 0 day neonate cerebellum cDNA, RIKEN full-length enriched
DE library, clone: C23009315 product: hypothetical protein, full insert
DE sequence (Mus musculus 12 days embryo spinal cord cDNA, RIKEN full-
DE length enriched library, clone: C530022L24 product: hypothetical
DE protein, full insert sequence).
GN Name=Pp1r15b;
OS Mus musculus (Mouse);
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
```

```
RA [1]
RC SEQUENCE FROM N.A.
RX STRAIN=C57BL/6J; TISSUE=Cerebellum, and Spinal cord;
RA MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA Carninci P., Hayashizaki Y.;
RA "High-efficiency full-length cDNA cloning.";
RA Meth. Enzymol. 303:19-44 (1999).
RN [2]
RC SEQUENCE FROM N.A.
RX STRAIN=C57BL/6J; TISSUE=Cerebellum, and Spinal cord;
RA MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA RIKEN FANTOM Consortium;
RA "Functional annotation of a full-length mouse cDNA collection.";
RA Nature 409:685-690 (2001).
RN [3]
RC SEQUENCE FROM N.A.
RX STRAIN=C57BL/6J; TISSUE=Cerebellum, and Spinal cord;
RA The FANTOM Consortium;
RA "Analysis of the mouse transcriptome based on functional annotation of
RA 60,770 full-length cDNAs.";
RA Nature 420:563-573 (2002).
RN [4]
RC SEQUENCE FROM N.A.
RX STRAIN=C57BL/6J; TISSUE=Cerebellum, and Spinal cord;
RA MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA Carninci P., Shibata Y., Hayatsu M., Sugahara Y., Shibata K., Itoh M.,
RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RA "Normalization and subtraction of cap-trapper-selected cDNAs to
RA prepare full-length cDNA libraries for rapid discovery of new genes.";
RA Genome Res. 10:1617-1630 (2000).
RN [5]
RC SEQUENCE FROM N.A.
RX STRAIN=C57BL/6J; TISSUE=Cerebellum, and Spinal cord;
RA MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA Konno H., Akiyama J., Nishi K., Kiteunai T., Tashiro H., Itoh M.,
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Fujiwaka S., Inoue K., Togawa Y., Izawa K., Tanaka T., Matsuura S., Kawai J.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RA "RIKEN integrated sequence analysis (RISA) system-384-format
RA sequencing pipeline with 384 multicapillary sequencer.";
RA Genome Res. 10:1757-1771 (2000).
RN [6]
RC SEQUENCE FROM N.A.
RX STRAIN=C57BL/6J; TISSUE=Cerebellum;
RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
RA Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,
RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
RA Katoh H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,
RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,
RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
RA Sasaki D., Shibata K., Shingawa A., Shiraki T., Sogabe Y., Tagami M.,
RA Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,
RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
RA Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
RN [7]
RC SEQUENCE FROM N.A.
RX STRAIN=C57BL/6J; TISSUE=Spinal cord;
RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
RA Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,
RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
RA Katoh H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,
RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,
RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
RA Sasaki D., Shibata K., Shingawa A., Shiraki T., Sogabe Y., Tagami M.,
RA Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,
RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
RA Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
```

```
RA Tomaru A., Tova T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
RL Submitted (APR-2002) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AK049028; BAC33517.1; -
DR EMBL; AK082957; BAC38708.1; -
DR MGI; MGI:244211; Ppplr15b.
DR GO; GO:0000164; C:protein phosphatase type 1 complex; IDA.
DR GO; GO:004722; F:protein serine/threonine phosphatase activity; IDA.
DR GO; GO:0045182; F:translation regulator activity; IC.
DR GO; GO:0006983; P:ER-overload response; IDA.
DR GO; GO:0006446; P:regulation of translational initiation; IC.
DR GO; GO:0042542; P:response to hydrogen peroxide; IMP.
DR GO; GO:0006979; P:response to oxidative stress; IDA.
KW Hypothetical protein.
SQ SEQUENCE 697 AA; 77711 MW; E439B12615F33737 CRC64;

Query Match 63.6%; Score 2419.5; DB 2; Length 697;
Best Local Similarity 68.9%; Pred. No. 1.4e-125;
Matches 482; Conservative 54; Mismatches 147; Indels 17; Gaps 6;

QY 1 MPEGTGGGRKRLGPRAGFRFPFPPPPRPSOAGSSKFPPLGPENSGNFTLLSSAQPETRV 60
DB 1 METGTHARKKPGPLGFWFLPFL-RRSHACSEFPFPPSSRQNGN-----SALPERRT 54

QY 61 SYWTKLLSLLAPLGLLQKVLMSQLFGCGMFTFRWLDFAGYISALRAKRGKPAAPTA 120
DB 55 RYWTKLLSLLALLPSLFQKLLWSQLFGGLIPTRWLDFAASYSALRALRGRESAAPT 114

QY 121 QKSLSSLOLDSDSPVTSFLDWLEGIHWQVSPDPLKLELKAKGSALDPAQAFLLEQOL 180
DB 115 QKSLSSLDSSDILVWSLWLEGLQWQSSDLELKAKQERLSDAAPTFLLEQOL 174

QY 181 WGVLLPSSLOSRLSYNRELSSPSGPNLQRIIDNFVSVVYLLNPSYLDLCPRLVSYON 240
DB 175 WGVLLPSSLAQLVSHRELSSSGPLSVQSLGNFKVSVYLLNPSYLDYLPQLGRQS 234

QY 241 SDGSEVVGFTLTPESSCLREDHCHPQLSAELIPASWQCGLPLSTGLPEIHLRMKR 300
DB 235 SAGGQGFVGRFTLTPESCYLSGDCGHPQLRAEWSATAWRRCPLSTGLPEIHLRMKR 294

QY 301 LEFLQANKGQDLPTPDQNGYHSLSEHSLLRMDPKHCRDNPQTFVPAAGDIPGNQTES 360
DB 295 LEFL-QANKGQELPTPDQNGYHSLSEHSLLRMDPKHCRDNPQTFVPAAGDIPGNQTES 349

QY 361 TEKIELLTTEVPLALEEESPSCPSSEIPEKEPGEGRISVVDYSYLEGDLPI SARPA 420
DB 350 TEKPELVIQEV-----SQSQGSLFCLEPVEKECEDHTNATDLSRGESLPVSTRPV 404

QY 421 CSNKLIDYLGAGSDLETSSDPEGEDWDEAEADDGFDSDSLSDLEQDPEGLHLWNS 480
DB 405 CSNKLIDYLGAGSDLEASDSESDGWERPEDDGFDSGLSESDVEQDSEGLHLWNS 464

QY 481 FCSVDYPNPQNTATIQTAAIRIVPEEPSDSSEKDLGKSDLENSSQSGSLPETPEHSSGEE 540
DB 465 FHSVDYPNPQNTATIQTAAIRAPRDPDSGTSWSGCVG-GSCQEGPLPETPDHSSGEE 523

QY 541 DWDESSADEASLKLWNSFCNSDDPNPLNFKAPFOTSGENEGKCRDSKTSSESVIAISE 600
DB 524 DWDEPSADEENLKLWNSFCHSEDPYNLLNFKAPFOPSGKWKWGQDSKASSEATVAFSG 583

QY 601 CHTLLSCVKQLGQSESCPSVQRDVLSGGRHTRHVKKVTFLFEVTEYIISGDEDRKG 660
DB 584 HHTLLSCKAQLLESQDNCPCGGLGEALAGERYTHIKRKVTFLFEVTEYIISGDEDRKG 643

QY 661 PWEEFARDGCFQKRIQETEDAIGYCLTFEHRERMFNRLO 700
DB 644 PWEEFARDGCFQKRIQETEAIGYCLAFEHRERMFNR 683

RESULT 4
Q658M2
ID Q658M2 PRELIMINARY; PRT; 407 AA.
AC Q658M2;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
```





QY 585 CRD--SKTPSESIVASECHTLLSKVQLLGSQ-----ESECPSVQRDVLSGGRH- 635  
 DB 496 AEDWGEAEPCCFRVAI-----YVPEKPPWAPPRLRLRLKRPETPTD 544  
 QY 636 -----VKRKVTFLVEVTEYIS-----GDERKGPWFSEFARDGCRFKRIQETDAIG 684  
 DB 545 PDPTPLKARKVRFSEKVTYVHFLAVWAGPAQAARQGFWEQLARDRSFARRITQABELS 604  
 QY 685 YCLTFEHRERMFNRLOQTCFKGLNLVKQ 712  
 DB 605 PCLTPAARAWARLNPPLAPIPALTQ 632

RESULT 9  
 ID 075807 PRELIMINARY; PRT; 674 AA.  
 AC 075807;  
 DT 01-NOV-1998 (TrEMBLrel. 08, Created)  
 DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)  
 DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)  
 DE Apoptosis associated protein (protein phosphatase 1, regulatory subunit 15A).  
 DE subunit 15A).  
 GN Name=GADD34; Synonyms=PPP1R15A;  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97298079; PubMed=9153226; DOI=10.1074/jbc.272.21.13731;  
 RA Hollander M.C., Zhan Q., Bae I., Fornace A.J. Jr.;  
 RT "Mammalian GADD34, an apoptosis- and DNA damage-inducible gene.";  
 RL J. Biol. Chem. 272:13731-13733(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TSSUS=Brain;  
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins L.F., Jordan H., Moore T., Max S.I., Wang J.J., Heish F.,  
 RA Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raba S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
 RA Krzywinski M.I., Skalska U., Small D.B., Schermer A., Schein J.E.,  
 RA Jones S.J., Marra M.A.;  
 RA "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences.";  
 RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC TSSUS=Brain;  
 RA Strausberg R.;  
 RL Submitted (Feb-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; U83981; AAC25631.1; --  
 DR EMBL; BC003067; AAH03067.1; --  
 DR Genew; HGNC:14375; PPP1R15A.  
 DR GO; GO:0006915; P:apoptosis; TAS.  
 DR GO; GO:0007050; P:cell cycle arrest; TAS.  
 DR GO; GO:0006974; P:response to DNA damage stimulus; TAS.  
 SQ SEQUENCE 674 AA; 73477 MW; B257AAL17456D1403 CRC64;

Query Match 6.8%; Score 260.5; DB 2; Length 674;  
 Best Local Similarity 21.9%; Pred. No. 2.8e-06;

Matches 164; Conservative 80; Mismatches 263; Indels 241; Gaps 32;  
 QY 70 LLAPLCLLQKVLWSQLFGMEPTTR-WLDFAGVYSAL--RALKGR----- 112  
 DB 21 LLSFVNLGSLRA--WSRL-RGLGFLPWLVEAVKGAALVEAGLEGEARTPLAHPWGR 77  
 QY 113 -----EKPAAPTAQKSLSLQDSDPSVTPDLWLEBEGIHWOYSPDILKLEKAKGSA 166  
 DB 78 RPBEAEADSGPGEDRETLGLKTSLSLPEAWGLLD-DDDGMGEREATSVY--RGQGSQ 133  
 QY 167 LDPAAQAFLEQLLWGLVGLLPSLSRLYSNRELSSPSGLPLNIQRIIDNFVSVYLLNPS 226  
 DB 134 FADGQRAPL-----SPSLIRTLQSDSKNPGKEAEAEAEAEAEAEAEAEAEAEAE 182  
 QY 227 YLDCFPRLVSYQNSDNGSEVGFQTLTPRESSCLREDH-CHPOPLSAELIPASWQCQCPPL 285  
 DB 183 HRECCPAVE--EEDD-----EEAVKEAHTSTSTALSFGSKPSTWVSCP-- 224  
 QY 286 STGLPEIHLRMKRLFLQOANKGQDLPTPDQDNGHSLSEHSLRLMDPKHCRNPTQ 345  
 DB 225 --GEEENQATEDKRTERSKGARK-----TSVSPRSSGSDPRS 259  
 QY 346 FVPAAGDIPGNTQESTEEKIELLTTEVPLALEBESPEGCPSSSE-----IPMEKEPG 397  
 DB 260 WYRSG-----EASEEKEEKAHKETGKEAAGPQSSAPAPORPOLKSWCQPSDEEG 312  
 QY 398 EGRISVVDYSLRGDLPIGARPACSNKLIDYILGGASSDLETSSDP----- 443  
 DB 313 EVK-----ALGAAEKDGEAECPCPPPPSAFLKAWV 344  
 QY 444 -EGEDWDEAEADGDFSDSLSDSLDLEQDDEG-----LHLWNSFCSDVPDYNQNF 492  
 DB 345 WPGEDTEEE-EDDEEDSDSGSDEBEAEASSTPATGVFLKSW-----VYQGED 396  
 QY 493 TATIQTAAIRVPEPSDEKLSKSDLENSSGSLP-----ETPEH-- 535  
 DB 397 T-----EEDEEDSDTGSADEREATSTPPASAFKAWVYRPGDEDEED 445  
 QY 536 ----SSGEEDWESSADEAESLWNSFCNSDDPYNPLNFKAPFQTSQ-----ENEKG 584  
 DB 446 EDVDSKEDDSEALGEAES-----DHPSPH-DQRAHFRGWYRPGKETEBEEA 495  
 QY 585 CRD--SKTPSESIVASECHTLLSKVQLLGSQ-----ESECPSVQRDVLSGGRH- 635  
 DB 496 AEDWGEAEPCCFRVAI-----YVPEKPPWAPPRLRLRLKRPETPTD 544  
 QY 636 -----VKRKVTFLVEVTEYIS-----GDERKGPWFSEFARDGCRFKRIQETDAIG 684  
 DB 545 PDPTPLKARKVRFSEKVTYVHFLAVWAGPAQAARQGFWEQLARDRSFARRITQABELS 604  
 QY 685 YCLTFEHRERMFNRLOQTCFKGLNLVKQ 712  
 DB 605 PCLTPAARAWARLNPPLAPIPALTQ 632

RESULT 10  
 ID 09NVU6 PRELIMINARY; PRT; 674 AA.  
 AC 09NVU6;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
 DE Hypothetical protein FLJ10499.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX PubMed=14702039; DOI=10.1038/ng1285;  
 RA Oka T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,  
 RA Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,  
 RA Sekine M., Obayashi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,



[illegible][illegible]



```
QY 232 PRLEVSQNSDGNSEVVGFTQTLTPSSCLREDHCHQPLSAELIPASWQCGPPLSTEGLP 291
DB 95 VEAESSPPETWGLSNVDYNA-KGQDDLRKEMERTAGKATLQAGLQAD-----145
QY 292 EIHILMRKRLFLFOQANKG---QDLPTPDQNGHSLRBEHSLR-----MDPKHR 340
DB 146 -----KRLGEVAREEGVAEPAYPT7SLEGGSPAENEDEGTVKTYQASAASIAPGYKP 198
QY 341 DNPTQFPAAGDIPGNTQESTEEKIELLTTTEVPLALEEESPSECPSS--EIPMEKEPGE 398
DB 199 STPVFFLGEA-----EHQATEEK-----GTE-NKADPSNPFSSGSHRAWYYSREKPKQ 247
QY 399 GRISVDVSYLEGDLPI SARPA-----CSNKLID-----YILGGAS 434
DB 248 -----EGEAKVRAHAGQGHCPCRNAEABEGGPETTFTVCTGNAFLKAWVYRPGED 296
QY 435 SDLETSSDPE-----CEDWDDEAEDDGFSDSLSDDL 468
DB 297 TEEEDNSDPSAEBDTAQTGATPHTSAFLKAWVYRPGEDTEEE-----DSDSDSAEEDT 350
QY 469 EQDPEGLHLWNSFCSDVPYNPONT-----ATIQTAARI---VPEEP 507
DB 351 AQTGATHT-SAFUKAWVYRPGEDTERENSDLDSAEEDTAQTGATPHTSAFLKAWVYRPG 409
QY 508 SDSEKOLSGKSDL-----ENSSQSGSLPET-----PEHSSGEEDDWES---545
DB 410 EDTEEE---NSDLDSAEEDTAQTGATPHTSAFLKAWVYRPGEDTDDTEEEEDSENVA PG 466
QY 546 ---SADBAESLKLWNSFC-----NSDDPYNPLNFKAPQTSGENEKGCRDSTPSES 594
DB 467 DSETADSSQSPCLQORCLPGEKTKGRGEEP--PL-FQVAFYLPGEK-----PES 513
QY 595 IVAISECHTLLSKVQLGS-----QSECEPDSVQDVLSCGRHHTVKKVTPLEEVTEY 650
DB 514 PWAAPKLPLRLQRLRUFKAPTRQDDPEIP-----LKKRVHFSEKVTVH 558
QY 651 YIS-----GDEDRKGPWEFARDGCRFQKRIQETEDAIGYCLTFEHRERMFNRQ 700
DB 559 FLAVWAGPAQAARRGPEQFARDRSFARRIAQAEKLGPLYLTDSRARAWRLR 613

RESULT 12
Q7TQC2 PRELIMINARY; PRT; 578 AA.
AC Q7TQC2;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DB Protein phosphatase regulatory subunit 15A.
GN Name=Gadd34;
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley;
RX MEDLINE=22697643; PubMed=12813455; DOI=10.1038/sj.onc.1206567;
RA Hollander M.C., Poole-Kella S., Fornace A.J. Jr.;
RT "Gadd34 functional domains involved in growth suppression and
RT apoptosis.";
RL Oncogene 22:3827-3832(2003).
DR EMBL; AY128642; AAM77795.1;
SQ SEQUENCE 578 AA; 63569 MW; 923EC49921C0BC61 CRC64;

Query Match 5.7%; Score 217; DB 2; Length 578;
Best Local Similarity 21.3%; Pred. No. 0.00058;
Matches 152; Conservative 80; Mismatches 225; Indels 258; Gaps 33;

QY 53 SAQPETRVSYTKLLS-QLLAPLGLLQKLVINSQLFGGMPPTWL--DFAGV-----102
DB 4 SPRPQ-HVLHWKEAHSFLLSPLMGFLSRA--WSRLRGPEVSEAWLAETVAGANQIAD 60
```

```
QY 103 -----YSALRAKGREKPA-APTAQKSUSLSQLDSSDPSTVTSPLDMLBEGIHQY 151
DB 61 LLTPPPVSENHPLURETEGNTPEWSKAAQRLCLDVEAQSSPKT-----WGL 108
QY 152 SPDLKLKELKAKSALDPAQAALLBQQLWGVELLFSSLSQSLRYSNRELSSSGPLNIQ 211
DB 109 SDID-----EHNGK-----PGQDGLREQVEVHTAGLP-TLQ-----PUHLQ 143
QY 212 RIDNPSVSYLLNPSYLDLCPRLVSVYQNSDGNSEVVGFTQTLTPSSCLREDHCHQPLS 271
DB 144 GADK-----KGVGVAREEGVSELA-----163
QY 272 AELIPAS-WQSCP---PLSTEGLEPEIHLRMKRLFLFOQANKGQDLPTPDQNGHYSLEE 327
DB 164 ---YPTSHWEGGPAEDEEDTETVKAH-----QASAASIAP-----GY-----198
QY 328 EHSLLRMDPKHCDNDTQFVPAAGDIPGNTQESTEEKI-----ELLTTTEVPLALE 377
DB 199 -----KPSTSVYVCPGEAHRATEEKGTDNKAEPSSGSHRWYVYHTRERPQEG 246
QY 378 EESPSGCPSSSETPMEKEPGEGRISVVVSYLYEGDLPI SARPAACSNKLI DYILGGASDL 437
DB 247 ETKPEHRAQOSHPQNAEAE-----EGGPETS---VCS-----GSAFL 282
QY 438 ETSSDPEGEDWDEAEDDGFSDSLSDSLQDQDPEGLHLWNSFCSDVPYNPONTATI Q 497
DB 283 KAWVYRPGEDTEEE-----EDSLDSAEED---TAHTCTTTPH-----TSAFL 321
QY 498 TAARI VPEEPSDEKDLGKSLSENSQSGSLPET-----PEHSSGEEDDWESSA 547
DB 322 KAWVYRPGEDTEEDDGDWDAAEEDASQCTTHTSAFLKAWVYRPGEDTEEDDSENVA 381
QY 548 D-BAESLKLWNSFCNSDDPYNPLNFKAPFOTSGENEGCRDSTPSESIV-----596
DB 382 PVDSETV---DSCQSTQHCLPV-----EKTGCGEAPFPFOVAFYLPQKQ PAP 427
QY 597 --AISECHTLLSKVQLL-----QSECEPDSVQDVLSCGRHHTVKKVTPLEEVTEY 650
DB 428 PWAAPKLPLRLQRLRSLFKAPARNQDPEIP-----LKKRVHFSEKVTVH 472
QY 651 YIS-----GDEDRKGPWEFARDGCRFQKRIQETEDAIGYCLTFEHRERMFNRQ 700
DB 473 FLAVWAGPAQAARRGPEQFARDRSFARRIAQAEKLGPLYLTAPFRARAWRLR 527

RESULT 13
Q6IN02 PRELIMINARY; PRT; 578 AA.
AC Q6IN02;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DB Myd116 protein.
GN Name=Myd116;
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Heart;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Straube R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shennen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buettow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heien F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
```





Job time : 81.7548 secs

---